

# Análise multivariada aplicada à caracterização de vinhos por LC-MS, com vista à criação de uma base de dados por região vitivinícola

**Inês Reis Marujo**

Mestrado em Química

Departamento de Química da Faculdade de Ciências  
da Universidade do Porto

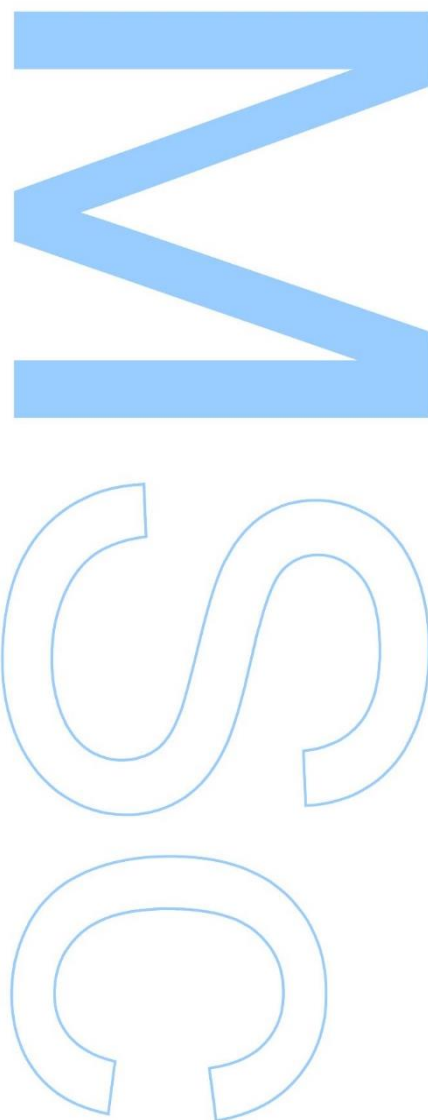
2016

## **Orientador**

Cristina Fernandes, Engenheira, Sogrape  
Vinhos

## **Coorientador**

Nuno Mateus, Professor Doutor, Faculdade  
de Ciências da Universidade do Porto

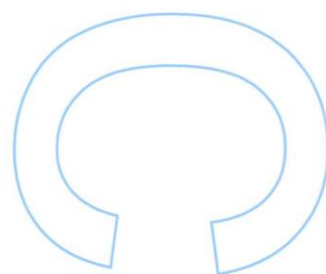
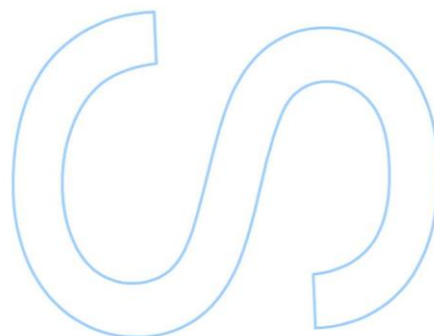
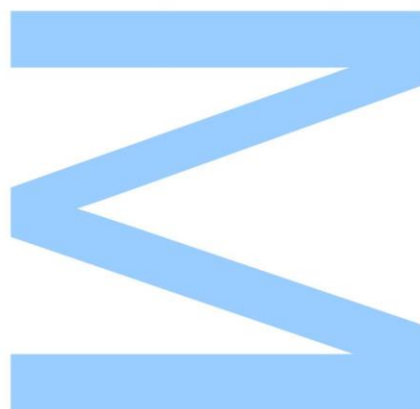




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, \_\_\_\_/\_\_\_\_/\_\_\_\_



## Agradecimentos

Gostaria de agradecer a todas as pessoas que, de alguma forma contribuíram para o bom desenvolvimento desde trabalho, ajudando-me a ultrapassar diversas dificuldades que foram acontecendo ao longo deste percurso:

Em primeiro lugar, gostaria de agradecer á Engenheira Cristina Fernandes pela sua disponibilidade, por todas as críticas construtivas, por todos os incentivos constantemente demonstrados, pela orientação, pelo apoio oferecido e também pela amizade demonstrada ao longo deste processo de crescimento profissional e pessoal que foi a elaboração deste relatório.

Ao Professor Doutor Nuno Mateus, o meu agradecimento pela constante disponibilidade e acompanhamento científico, bem como todos os apoios demonstrados para que este projeto corresse sempre da melhor forma, mostrando uma elevada compreensão e incentivo em todas as adversidades que foram surgindo.

Gostaria de dirigir os meus sinceros agradecimentos a todos os elementos da empresa Sogrape Vinhos S.A., que me acolheram durante o período de estágio e que a todos os níveis muitos ensinamentos me transmitiram.

Ao Engenheiro Nuno Neves, o meu profundo agradecimento pelas valiosas discussões científicas e apoios sobre espetrometria de massa e PCA, por toda a paciência e compreensão, por todo o incentivo, companheirismo e amizade sempre demostradas ao longo deste projeto.

Aos meus pais, irmã e avós por todos os apoios sempre demostrados, por toda a compreensão e carinho que me deram ao longo destes anos, fazendo-me sempre acreditar que iria ser possível concretizar este sonho por forma a crescer de forma pessoal e profissional.

A todos, o meu muito obrigada.

## Resumo

A química do vinho é bastante complexa e é através da análise química com o recurso a técnicas espectroscópicas que é possível detetar fraudes e compreender os processos que ocorrem no fabrico do vinho. Com a mesma velocidade de evolução das tecnologias, as fraudes e as formas com que elas acontecem, vêm crescendo de forma considerável na sociedade atual atingido mundialmente as empresas produtoras de vinho, ocasionando perdas substanciais nos respetivos patrimónios.

Este projeto decorreu na Sogrape Vinhos S.A, e tem como objetivo a criação de uma base de dados de vinhos por região vitivinícola, em que se pretende identificar a proveniência dos vinhos comprados, detetar fraudes e contaminações acidentais. As principais tarefas a desenvolver durante o estágio são desenhar o protocolo para a criação da base de dados e testá-lo, sendo estas as maiores necessidades atuais da empresa.

A primeira parte deste trabalho incidiu numa pesquisa bibliográfica exaustiva de forma a conhecer melhor as metodologias usadas na análise de vinhos bem como a sua composição química, os princípios de funcionamento do espectrómetro de massa acoplado com cromatografia líquida (LC-MS) e trabalhos que foram feitos em Portugal, França e Itália que visam à criação de uma base de dados, bem como os procedimentos adotados.

Cada amostra que chega ao laboratório é identificada pelo seu tipo, ano e região, sendo estas posteriormente preparadas e injetadas no LC-MS. O tratamento de resultados é feito recorrendo a ferramentas estatísticas, usando o programa ProfileAnalysis que tem por base a análise de componentes principais (PCA).

Observa-se uma boa separação entre as diferentes regiões vinícolas em estudo, não só pelo seu ano, mas também pela sua casta.

**Palavras-chave:** Vinho, castas, região, base de dados, LC-MS, PCA

## Abstract

Wine chemistry is quite complex, frauds can be detected through chemical analysis and spectroscopy techniques in a way that all the processes that occur in wine production can be understood. As technology evolves, so do the frauds which translates on substantial losses to wine-producing companies.

This project took place in Sogrape S.A., with aim to create a wine region database, making possible to identify the wine origin, detect fraud and accidental contamination. The main tasks during the trainship are the protocol design to create a database and test it, currently these two are the greatest needs of the company.

The first part of this work focused on a comprehensive literature search in order to better understand the methods used in the wines analysis as well as their chemical composition, the operating principles of mass spectrometer coupled with liquid chromatography (LC-MS) and works made in Portugal, France and Italy in order to create a database, as well as the procedures adopted.

Reaching the laboratory, each sample is identified by its type, year and region, and then prepared and injected into the LC-MS. Results treatment were done by the help of statistical tools, using a ProfileAnalysis program based on a principal component analysis (PCA).

In this study a good separation between wine of different regions was achieved, not only for its year, but also for its variety.

**Keywords:** Wine, caste, region, database, LC-MS, PCA

## Índice geral

Agradecimentos.....	iii
Resumo .....	iv
Abstract.....	v
Índice de figuras .....	ix
Índice de tabelas.....	xi
Notação e glossário.....	xii
Preâmbulo .....	xiii
INTRODUÇÃO GERAL .....	1
1. O vinho .....	2
2. Caracterização das principais regiões vitivinícolas de Portugal.....	3
3. Castas .....	5
Características de algumas castas plantadas em Portugal .....	5
4. Fraudes.....	7
5. Espectrometria de massa acoplada com cromatografia líquida .....	8
HPLC .....	8
MS .....	8
Vácuo.....	10
5.1. Método de ionização.....	10
Eletrospray.....	11
5.2. Método de separação (analisador de massa).....	12
Quadrupolo.....	12
TOF .....	14
5.3. Detetor .....	16
Multiplicador de eletrões .....	16
5.4. Modos de aquisição .....	17

6. Análise de componentes principais .....	18
6.1. Análise multivariada.....	18
6.2. PCA .....	19
Explicação do processo .....	20
Vantagens e desvantagens do método PCA .....	23
DESCRIÇÃO DO TRABALHO.....	25
1. Apresentação e enquadramento do projeto .....	26
2. Apresentação e caracterização da instituição acolhedora de estágio .....	26
3. Organização da tese .....	27
MATERIAIS E MÉTODOS .....	28
1. Amostras e processo de amostragem .....	29
2. Reagentes .....	31
3. Equipamento .....	31
4. Preparação da amostra .....	31
5. Análise por LC-MS .....	33
Método HPLC .....	33
Método MS .....	34
Calibrações .....	35
RESULTADOS E DISCUSSÃO .....	37
1. Calibração dos espectros de LC-MS .....	38
2. PCA .....	39
Amostras de 2013, quando ionizadas em modo negativo .....	40
Diferenciação das castas em amostras de 2013 quando ionizadas em modo positivo .....	44
Comparação das amostras de 2013 quando ionizadas em modo negativo e em modo positivo .....	44
Diferenciação das amostras por ano 2013, 2014 e 2015 – modo positivo .....	47
CONCLUSÕES .....	50
Considerações finais e perspetivas futuras.....	52

REFERÊNCIAS BIBLIOGRÁFICAS.....	53
ANEXOS .....	57
Anexo 1 – Tabela dos <i>buckets</i> com maior variância do ano de 2013 .....	57



## Índice de figuras

Fig.1 – Regiões vitivinícolas de Portugal .....	4
Fig.2 - Autosampler do LC .....	8
Fig.3 – Espectrómetro de massa acoplado com cromatógrafo líquido (LC-MS).....	9
Fig.4 – Processo de análise em espectrometria de massa.....	9
Fig.5 – Demonstração do funcionamento da fonte de ionização – ESI.....	11
Fig.6 – a) Corrente contínua aplicada quadrupolo; b) Corrente alternada aplicada no quadrupolo ....	12
Fig.7 – Demonstração do funcionamento do quadrupolo .....	13
Fig.8 – Analisador de massa: hexapolo, quadrupolo e célula de colisão. ....	14
Fig.9 – Princípio de operação do analisador TOF-MS .....	15
Fig.10 – Princípio de operação do analisador TOF-MS - acelerador ortogonal.....	15
Fig.11 – Princípio de operação do analisador TOF-MS - refletor.....	16
Fig.12 – Princípio de operação do multiplicador de eletrões.....	17
Fig.13 – Digitalizador.....	17
Fig.14 – Exemplificação de um espectro obtido por LC-MS .....	18
Fig.15 – Relação entre as variáveis das diferentes componentes principais .....	19
Fig.16 – Matriz de dados na análise por PCA .....	20
Fig.17 – Combinações lineares entre as componentes principais e as variáveis originais da matriz X, na análise por PCA.....	20
Fig.18 – Matriz covariância das novas variáveis na análise por PCA .....	21
Fig.19 – Demonstração do critério de Pearson .....	22
Fig.20 – Demonstração do critério de Kaiser-Guttman .....	22
Fig.21 – Exemplo de um gráfico obtido pelo critério de <i>Scree Plot</i> ou teste de <i>Cattell</i> .....	22
Fig.22 – Mapa das regiões vinícolas portuguesas em estudo (assinaladas no respetivo mapa).....	29
Fig.23 – Chegada das amostras ao laboratório .....	32
Fig.24 – Selagem dos frascos antes da análise por LC-MS .....	33

Fig.25 – Calibração da amostra através da válvula de 6 portas .....	35
Fig.26 – Sobreposição de quatro cromatogramas (BPC) ao fim de calibrados, modo positivo .....	39
Fig.27 – <i>Bucket table</i> calculada para as amostras de 2013, em modo negativo .....	40
Fig.28 – Seleção do método de calibração a ser usado .....	41
Fig.29 – Escolha das componentes principais de forma visual.....	41
Fig. 30 – Análise da distância das amostras ao modelo .....	42
Fig. 31 – Semelhança das amostras entre si, e a distância de cada uma delas ao modelo (ionização em modo negativo das amostras de 2013).....	43
Fig. 32 – Variação da intensidade do <i>bucket</i> selecionado em todas as amostras.....	43
Fig. 33 – Diferenciação das castas em amostras de 2013 .....	44
Fig. 34 – Escolha das componentes principais quando as amostras de 2013 foram ionizadas em modo positivo.....	45
Fig. 35 – Comparação da variância dependendo do modo de ionização. a) ionização positiva; b) ionização negativa .....	46
Fig. 36 – Diferenciação entre as amostras de 2013 e 2014.....	47
Fig. 37 – Análise das diferentes regiões das amostras de 2013, 2014 e 2015 .....	48
Fig. 38 – Análise dos diferentes anos das amostras 2013, 2014 e 2015 .....	48

## Índice de tabelas

Tabela 1: Número de amostras analisadas no ano de 2013 nas diferentes regiões em estudo .....	30
Tabela 2: Número de amostras analisadas no ano de 2014 nas diferentes regiões em estudo .....	30
Tabela 3: Número de amostras analisadas no ano de 2015 nas diferentes regiões em estudo .....	30
Tabela 4: Exemplo da identificação de uma amostra após a sua chegada ao laboratório .....	32
Tabela 5: Matriz de códigos usada para uma melhor identificação das diversas amostras .....	32
Tabela 6: Descrição dos constituintes dos solventes A e B, utilizados na fase móvel do HPLC .....	33
Tabela 7: Percentagem de cada um dos solventes usados na fase móvel ao longo do tempo de aquisição com um fluxo de 0,3mL/min .....	34
Tabela 8: Principais condições de análise no método MS .....	34
Tabela 9: Propriedades químicas dos calibrantes e dos padrões internos usados em LC-MS.....	36
Tabela 10: Condições usadas no programa <i>DataAnalysis</i> para a calibração dos espetros.....	38
Tabela 11: Condições usadas no programa <i>ProfileAnalysis</i> .....	39
Tabela 12 – Exemplos de possíveis estruturas para os <i>buckets</i> com maior variância.....	46

## Notação e glossário

<b>AC</b>	Corrente alternada
<b>ASAE</b>	Autoridade de Segurança Alimentar e Económica
<b>CheBI</b>	Chemical Entities of Biological Interest
<b>CP</b>	Componente Principal
<b>Da</b>	Dalton
<b>DC</b>	Corrente contínua
<b>DIV</b>	Mixture of unknown varieties
<b>EN</b>	Encruzado
<b>ESI</b>	Eletrospray
<b>FR</b>	Tinta Francisca
<b>UHPLC</b>	Ultra high-performance liquid chromatography
<b>KEGG</b>	Kyoto Encyclopedia of Genes and Genomes
<b>LC-MS</b>	Liquid chromatography–mass spectrometry
<b>m/z</b>	Razão massa/carga
<b>MS</b>	Mass spectroscopy
<b>OIV</b>	Organização Internacional da Vinha e do Vinho
<b>PCA</b>	Principal Component Analysis
<b>RZ</b>	Tinta Roriz
<b>TC</b>	Tinto Cão
<b>TN</b>	Touriga Nacional
<b>TF</b>	Touriga Franca
<b>TOF</b>	Time-of-flight

## Preâmbulo

Na introdução geral é apresentada uma breve caracterização das diferentes regiões vitivinícolas portuguesas e é feita uma pequena referência a alguns casos de fraudes em vinhos. São descritos os princípios do método de análise – LC-MS – e do método do tratamento dos dados – PCA.

É apresentado o enquadramento do trabalho e os motivos pelos quais é importante desenvolver bases de dados, para ser possível identificar fraudes e outros tipos de contaminações no vinho.

De seguida, são apresentados os reagentes e equipamento utilizados, bem como o procedimento adotado na preparação da amostra assim como as condições utilizadas na análise por LC-MS.

Na secção correspondente aos resultados e discussão são apresentados espectros originais obtidos bem como a otimização das condições no tratamento dos dados por PCA, recorrendo ao programa ProfileAnalysis e os resultados por este obtidos.

Seguidamente, às conclusões obtidas é feita uma avaliação do trabalho realizado, principalmente no que diz respeito aos objetivos cumpridos e às principais limitações.

## **Capítulo I**

### **INTRODUÇÃO GERAL**

Neste capítulo, é feita uma breve caracterização do vinho e das diferentes regiões vitivinícolas portuguesas a nível do clima, solo e até de algumas castas. É ainda feita uma pequena referência a alguns casos conhecidos de fraudes em vinhos.

De seguida, são apresentadas noções base sobre os princípios do método de análise utilizado, o LC-MS, bem como do método usado no tratamento de dados, o PCA.

## 1. O vinho

O vinho é uma das bebidas mais consumidas do mundo, o que faz com que a sua produção e exportação tenha aumentado <sup>[4]</sup>. Tal facto, implica a necessidade de um rigoroso controlo na qualidade e segurança alimentar deste tão prestigiado produto. Com os inúmeros progressos que tem sofrido, tornou-se alvo de um grande interesse científico que visa o conhecimento das inúmeras variáveis que alteram o processo de vinificação. Neste sentido, as melhores ferramentas para assegurar a conformidade e qualidade do produto são as análises físicas, químicas, microbiológicas e sensoriais, que permitem a identificação rápida e atempada de defeitos ou problemas e auxiliam na melhoria e manutenção da qualidade dos vinhos <sup>[36]</sup>.

Obtido pela vinificação/fermentação alcoólica, total ou parcial de uvas frescas, provenientes de vários tipos de castas (*Vitis Vinifera*), o vinho é composto por diversas substâncias químicas que possuem características próprias, de concentrações variadas e que se combinam entre si, originando um produto alimentar de elevada complexidade <sup>[15]</sup>.

A composição química exata das uvas depende da casta utilizada e de diversos fatores agronómicos e ambientais que incluem o tipo de solo, a topografia do local e as condições atmosféricas <sup>[5]</sup>. Como tal, as análises físico-químicas, representam um importante suporte para o acompanhamento da vinificação, de forma a aumentar a qualidade do produto, reduzir os desperdícios e conservar o vinho por um período de tempo mais extenso <sup>[2,4]</sup>.

Os objetivos da realização das análises rotineiras são diversos, como determinar quantidades de alguns componentes de forma a verificar se estão dentro do limite legal ou avaliar a qualidade geral do vinho, mas, nem todas as análises têm a capacidade/intuito de detetar falsificações, fraudes ou detetar a presença de produtos proibidos <sup>[5,36]</sup>. Como tal, há a necessidade de se recorrer a técnicas mais avançadas como é o caso da espetrometria de massa acoplada com cromatografia líquida para ser possível chegar a conclusões mais específicas, como é o caso deste trabalho, que tem como objetivo criar uma base de dados de vinhos por região vitivinícola, de forma a identificar a proveniência dos vinhos comprados, detetar fraudes e contaminações acidentais.

## 2. Caracterização das principais regiões vitivinícolas de Portugal

[3,23,30,40]

Portugal é um excelente país produtor de vinho, devido ao seu clima ameno e à biodiversidade da vinha, quando comparado com outros países ocidentais. O cultivo da vinha está praticamente disseminado por todo o território português, estando o mesmo dividido em diversas regiões vitivinícolas, como se pode observar na figura 1, sendo o vinho produzido nessas regiões naturalmente diferente.

De entre os vinhos portugueses, aquele que mais sucesso conheceu em Portugal foi o Vinho do Porto. O Vinho do Porto é uma bebida licorosa produzida na Região Demarcada do Douro. Este vinho apresenta uma riqueza e intensidade de aromas, um teor alcoólico elevado bem como uma vasta gama de doçuras e uma grande diversidade de cores. O Vinho do Porto é uma imagem de marca de Portugal e serviu de pilar para cimentar a imagem dos vinhos portugueses além-fronteiras. A primeira região demarcada e regulada no mundo, o Alto Douro Vinhateiro, é um exemplo excecional de uma região tradicional que vive em torno da produção vinícola, seja do afamado vinho do Porto, seja dos mais recentes vinhos de mesa de grande qualidade. A região do Douro apresenta uma forte inclinação do terreno e os seus solos são compostos por xisto e granito. Relativamente ao clima este é habitualmente seco, com invernos frios e verões muito quentes. As principais castas são a Touriga Franca, Tinta Roriz, Tinta Barroca, Touriga Nacional, Tinta Amarela, Tinto Cão, Sousão.

A Região Demarcada dos Vinhos Verdes, é tradicionalmente conhecida como a zona de Entre-Douro-e-Minho, sendo que a sua área geográfica é a maior Região Demarcada Portuguesa, e uma das maiores da Europa. As condições naturais desta região são as ideais para a produção de excelentes vinhos brancos, assim como espumantes e aguardentes. A região dos vinhos verdes apresenta um nível de humidade atmosférica relativamente alta o que provoca níveis de precipitação elevados. É rica em recursos hidrográficos e os seus solos são maioritariamente graníticos, variando entre férteis a muito férteis, de acidez elevada. O seu clima é influenciado pelas brisas marítimas o que faz com que a temperatura nesta região seja amena durante todo o ano. A região é caracterizada por produzir grandes vinhos que, pela sua frescura natural e baixo teor alcoólico, mostram enorme potencial para produção de bons espumantes. Já os teores em acidez natural dos bagaços, bem como as suas características organoléticas, mostram condições técnicas excelentes para a produção de grandes aguardentes bagaceiras, a partir da destilação dos bagaços, e de excelentes aguardentes vínicas, fruto da destilação dos vinhos. As principais castas são, para os brancos, o Loureiro, o Alvarinho, o Arinto (conhecido localmente por Pedernã) e a Trajadura. Para os tintos são o Vinhão e para rosados o Espadeiro.



A zona do Dão situa-se na região da Beira Alta, no centro Norte de Portugal nas serras do Caramulo, Montemuro, Buçaco e Estrela. As condições geográficas são excelentes para produção de vinhos. A região é extremamente montanhosa, contudo a altitude na zona sul é menos elevada e os seus solos são essencialmente xistosos e graníticos. O clima nesta região sofre influência tanto do atlântico como das regiões do interior o que provoca que os invernos sejam frios e chuvosos e os verões são quentes e secos. As principais castas usadas para o fabrico de vinho nesta região são o Alfrocheiro, Aragonês, Bastardo, Jaen, Rufete, Tinto-Cão, Touriga Nacional e Trincadeira.

O vasto e diferenciado território do Alentejo encontra-se dividido administrativamente em três distritos, Portalegre, Évora e Beja que, juntos, perfazem as fronteiras naturais do Vinho Regional Alentejano. Apresenta solos litólicos não húmicos, pobres em matéria orgânica, e a capacidade de retenção de água varia com a estrutura dos solos. Relativamente ao clima, é mediterrânico, quente e seco, com influência continental. No Alentejo, para além das muitas castas autóctones que imprimem um forte carácter regional, variedades perfeitamente adaptadas à geografia e às condicionantes da paisagem alentejana, primam outras variedades forâneas de introdução relativamente recente, castas de valor reconhecido que reforçam a liderança vitivinícola do Alentejo, como é o caso do Aragonês, Castelão, Trincadeira, Alfrocheiro, Alicante Bouschet, Cabernet Sauvignon, Carignan, Grand Noir, Moreto e Tinta Caiada, Cinsaut, Corropio, Tinta Carvalha.



Fig.1 – Regiões vitivinícolas de Portugal

### 3. Castas

A identificação de castas é importante para os viticultores, produtores de vinho, autoridades reguladoras e consumidores. O aumento do interesse na sua identificação resultou de uma necessidade de plantar e produzir uvas e produtos derivados corretamente identificados, devido ao valor económico em conformidade com designações de origem protegida. A correta identificação de castas da videira é crucial, e deste modo é necessário desenvolver métodos eficientes de identificação da proveniência da casta e também durante o processo de produção de vinho.

- **Características de algumas castas plantadas em Portugal**

A Touriga Franca é uma das castas mais plantadas na zona do Douro e Trás-os-Montes. É considerada umas das melhores castas para a produção de vinho do Porto e do Douro, tem produções regulares ao longo do ano e é bastante resistente a doenças. Os seus cachos são médios ou grandes com bagos médios e arredondados. Os vinhos produzidos por esta casta têm uma cor intensa e são bastante frutados. No vinho do Porto, a Touriga Franca integra os lotes com a Tinta Roriz e a Touriga Nacional <sup>[29]</sup>.

A Touriga Nacional é uma casta nobre e muito apreciada em Portugal. Inicialmente cultivada na região do Dão, rapidamente foi expandida à zona do Douro para ser utilizada na produção de vinho do Porto. É uma casta de pouca produção, possui cachos abundantes, mas pequenos. Os bagos têm uma elevada concentração de açúcar, cor e aromas. Os vinhos produzidos ou misturados com a casta Touriga Nacional são bastante equilibrados e com boa capacidade de envelhecimento <sup>[22,29]</sup>.

A casta Tinto Cão é cultivada na zona do Douro desde o século XVIII, contudo como era pouco produtiva nunca foi muito apreciada pelos agricultores. Possui cachos muito pequenos e de maturação tardia. É muito resistente a doenças e à podridão, além de suportar temperaturas muito elevadas. Produz vinhos de carregados de cor e de aromas delicados e florais <sup>[22,29]</sup>.

A casta Tinta Francisca é uma casta sem grande implantação utilizada normalmente na produção de vinhos de lote e é plantada quase exclusivamente na região do Douro. Casta pouco produtiva de cacho pequeno com bagos arredondados de tamanho médio e de cor negro-azul <sup>[22]</sup>.

O cultivo da casta Encruzado é praticamente exclusivo da zona do Dão, sendo provavelmente a melhor casta branca plantada na região. É utilizada na produção da maioria dos vinhos brancos através de vinhos de lote ou de vinhos monovariais. Tem uma boa produção, sendo bastante equilibrada em açúcar e acidez. Por outro lado, é muito sensível à podridão e a condições climáticas desfavoráveis (chuva e vento). Os vinhos compostos por esta casta são muito aromáticos e de sabor acentuado. Apresentam uma longevidade fora do comum, uma vez que podem conservar-se em garrafa durante muitos anos <sup>[22,29]</sup>.

A Tinta Roriz é uma casta muito adaptável a diferentes climas e solos, por isso o seu cultivo tem aumentado e alargado para as regiões do Dão, Ribatejo e Estremadura. Para as características da casta serem excelentes, a sua produção tem de ser controlada sendo que as condições ideais são solos arenosos e argilo-calcários em climas quentes e secos, para que a produção seja menor e os bagos mais concentrados. Esta casta origina vinhos de elevado teor alcoólico, de baixa acidez e indicados para envelhecer, sendo muito resistentes à oxidação <sup>[29]</sup>.

## 4. Fraudes

As fraudes em vinhos são divididas em 3 grandes modalidades: as fraudes por alteração enzimáticas, química e microbiológica, por adulteração com a adição ou subtração de algum composto ilegal, por falsificação e por sofisticação <sup>[15]</sup>. Do mesmo modo que as técnicas fraudulentas evoluem, os sistemas de deteção das mesmas também têm de continuar a melhorar. A maioria das fraudes é de difícil perceção para os consumidores, então cabe às entidades responsáveis pela produção do vinho verificar e controlar os possíveis tipos de fraudes que podem ocorrer, de forma a garantir uma elevada qualidade do produto <sup>[31,33]</sup>.

Um dos casos que envolveu a falsificação de vinhos, tendo em conta a sua região de produção teve origem em Portugal e Espanha. A notícia foi avançada pelo Diário de Notícias em que “Vinho espanhol usado para falsificar marcas portuguesas”. O vinho espanhol é menos denso e mais claro do que o português, ou seja, alguns falsificadores aproveitam “as nossas uvas carregadas de cor e misturam-nas para incrementar o volume”. Esta fraude é a mais detetada pela ASAE e envolve grandes armazenistas.

As principais fraudes ao consumidor estão relacionadas com a rotulagem, com informação fraudulenta sobre o estágio de barrica e com uma vinificação manipulada. Nesta as mais conhecidas são a adição de água, edulcorantes e corantes sintéticos, o uso de castas “*non-Vitis vinífera*” e a adição de glicerol ao vinho. Existe também a possibilidade de ocorrerem contaminações acidentais no transporte do vinho, em que este seja transportado em camiões mal lavados e que anteriormente tenham transportado detergentes, leite ou até mesmo óleos lubrificantes. No que diz respeito às fraudes que as empresas podem sofrer estão relacionadas com a compra de vinho biológico em que um fator importante em ter em consideração é a pesquisa de pesticidas. É através da análise química que é possível detetar a maior parte das fraudes no fabrico do vinho e como tal, é necessário investigar mais sobre esta temática que tanto preocupa os produtores e os consumidores <sup>[31,32]</sup>.

## 5. Espetrometria de massa acoplada com cromatografia líquida

O LC-MS é uma técnica que combina a capacidade de separação do HPLC com o poder de deteção do espectrómetro de massa. É atualmente o método aceite para traçar perfis metabólicos do vinho sendo rotineiramente usado para analisar a quimiodiversidade do vinho e dos seus benefícios. Esta aplicação providencia uma técnica de elevado potencial para o controlo, qualidade e autenticidade do vinho [8,17,19,20,21,24].

- **UHPLC**

O UHPLC permite a separação de compostos de uma mistura com base no carácter hidrofóbico dos compostos [9].

A introdução da amostra líquida no LC é feita recorrendo à utilização de um *auto sampler*, é ótimo pois permite analisar uma sequência de amostras sem intervenção do utilizador, assegurando que a injeção é feita sempre da mesma forma, permitindo a automação de misturas. O solvente (fase móvel) é bombeado de forma a arrastar a amostra através da coluna (fase estacionária) onde ocorre a separação dos compostos por afinidade à coluna.



Fig.2 - Autosampler do LC

- **MS** [17,20,24,26]

A espetrometria de massa é utilizada para detetar, identificar e quantificar as moléculas de uma amostra depois de serem convertidas a iões com base na sua razão massa/carga ( $m/z$ ) produzindo sinais resultantes da abundância de cada espécie iónica presente. Tem a capacidade de obter informação qualitativa (estrutura) e quantitativa (concentração) nas moléculas de um composto, podendo produzir e detetar fragmentos de moléculas. Ou seja, o conceito é formar iões na amostra, separá-los de acordo com a sua

razão massa/carga e quantificá-los. A espetrometria de massa acoplada com cromatografia líquida é uma técnica sensível, robusta e uma ferramenta indispensável no estudo de pequenas quantidades de amostra (baixa concentração), sendo capaz de captar e analisar pequenas ou grandes moléculas de várias polaridades numa matriz biológica complexa.

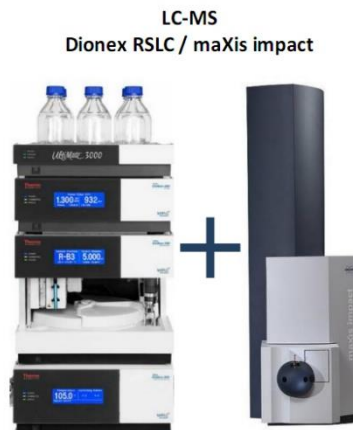


Fig.3 – Espetrómetro de massa acoplado com cromatógrafo líquido (LC-MS), equipamento utilizado.

O ambiente de análise de espetrometria de massa é todo controlado em computador (ionização, deteção e aquisição). As moléculas gasosas são ionizadas e fragmentadas numa fonte de ionização. Todos os iões com diferentes razões  $m/z$  passam por um analisador de massa sendo depois detetadas. Chegados ao detetor, os iões são convertidos num sinal elétrico e posteriormente num sinal digital, obtendo-se um espetro de massa, em que a intensidade dos iões corresponde à quantidade presente [8, 19, 24].

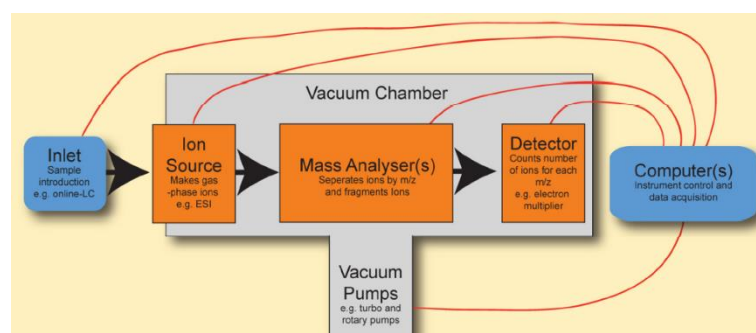


Fig.4 – Processo de análise em espetrometria de massa.

- **Vácuo**

O primeiro requisito para um bom desempenho do MS é o desempenho em alto vácuo. Se tal não acontecesse, os iões do analito ionizado não resistiam à colisão com as moléculas de ar no analisador<sup>[24]</sup>. A separação dos iões em sistema de vácuo é fundamental para estes não colidirem durante o processo. Se ocorrem colisões entre moléculas, o equipamento sofre uma redução na resolução e na sua sensibilidade. É fundamental assegurar que os iões estão em fase gasosa antes de serem separados e detetados.<sup>[24,26]</sup>

### 5.1. Método de ionização

A fonte de ionização é a parte do equipamento responsável por converter os analitos de interesse em iões em fase gasosa, sendo um pré-requisito fundamental para qualquer análise por MS. A compreensão dos mecanismos envolvidos na ionização é de extrema importância visto que, uma das limitações do ESI é apenas permitir a análise de compostos que sejam ionizáveis dependendo da diferença de potencial aplicada serão obtidos no modo positivo ou negativo<sup>[17,24,26]</sup>.

A sensibilidade da resposta obtida recorrendo à MS depende do tipo de interface que é utilizada. As interfaces de ionização mais utilizadas correspondem a fontes de ionização a pressão atmosférica por *eletrospray* (ESI) e por ionização química<sup>[21]</sup>.

As moléculas são protonadas no modo de ionização positivo e desprotonadas no modo de ionização negativo, podendo ocorrer a formação de iões adutos que vão surgir no espetro de massa. Mas, por sua vez o uso combinado dos modos de ionização positivo e negativo pode também facilitar a identificação de compostos em estudo<sup>[24]</sup>.

A escolha do tipo de ionização depende das propriedades do analito (estabilidade térmica, volatilidade, polaridade, etc), mas todos eles são baseados na razão  $m/z$ , velocidade de análise, transmissão, exatidão, resolução e gama de massas<sup>[17,19]</sup>.

O processo onde as moléculas eletricamente carregadas, ganham ou perdem um ou mais eletrões é chamado de ionização, havendo um mínimo de energia, chamado de “potencial de ionização”, que providencia a ordem para que a formação dos iões ocorra. Existe a criação de cargas positivas e negativas, dependendo da afinidade protónica da amostra, sendo que a perda de iões é inevitável<sup>[26]</sup>.

## • Eletrospray

A interface entre o UHPLC e o espectrómetro de massa é crítico para o desempenho do sistema, apesar de todos eles serem concebidos para transferir amostras ionizadas. Na maioria dos sistemas de LC-MS a configuração dessa interface pode ser modificada para produzir iões, com cargas positivas ou negativas [21,26].

A fonte de ionização mais utilizada para o acoplamento LC-MS é a de *electrospray* (ESI), descrita inicialmente por *Fenn*. Basicamente, os analitos em solução são ionizados quando atravessam um capilar metálico, onde uma voltagem é aplicada. O líquido (amostra) que contém o analito é sujeito a um potencial elevado, formando-se gotículas carregadas, sendo que à medida que atravessam um gradiente de elevada pressão, as gotas tornam-se mais pequenas [19,21]. Em geral, no caso de analitos básicos, é adicionado um aditivo à fase móvel com o objetivo de diminuir o arrastamento dos picos comatográficos (*tailing*). Esses aditivos devem ser voláteis e presentes em baixas concentrações, afim de se evitar que o mesmo interfira no processo de ionização do analito.

O líquido atravessa o cone de *Taylor* formando gotas mais pequenas. A pulverização é feita aplicando um gás de nebulização com alto fluxo e baixa temperatura (azoto), levando à formação de iões em fase gasosa, que contêm além da fase móvel, iões carregados positivamente em excesso [21]. Esse gás tem como objetivo vaporizar a amostra, diminuindo as gotas previamente formadas até o limite onde a repulsão entre as cargas positivas, localizadas preferencialmente na superfície das mesmas, supere a força de coesão dessa gota (tensão superficial). Nesse momento, cargas positivas são ejetadas da gota, levando a formação dos iões positivos de interesse. No caso de análise em modo negativo, o fenómeno ocorre de modo semelhante, mas as polaridades são alteradas. Uma das funções da interface é ionizar os compostos-alvo, uma vez que muitos dos componentes do eluente do UHPLC não são carregados. [17,26]. O líquido é pulverizado para fora da extremidade do tubo e as gotículas carregadas diminuem rapidamente de tamanho.

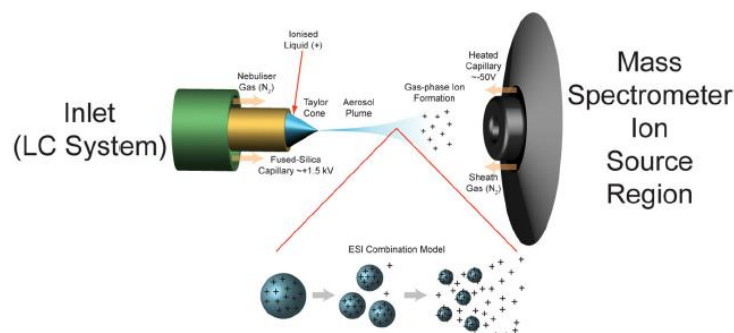


Fig.5 – Demonstração do funcionamento da fonte de ionização – ESI.



O *eletrospray* é recomendado para compostos ionizáveis, sendo que é uma técnica que resulta em pouca fragmentação. Contudo, é muito propícia a contaminações e as moléculas com muitos eletrões de valência contribuem para uma grande complexidade espectral [19, 24].

## 5.2. Método de separação (analisador de massa)

Uma vez que a amostra é ionizada, deve ser orientada para a parte do analisador do espectrómetro de massa, com uma lente de focagem eletricamente carregada com o mesmo tipo de carga dos iões para concentrá-los num feixe [20].

O coração do espectrómetro de massa é o analisador, pois separa os iões pela sua razão de  $m/z$ . Utiliza os campos elétricos ou magnéticos ou uma combinação de ambos para deslocar os iões de região em que são produzidos para o detetor [24].

*Tandem MS* ou *MS/MS* é um modo de análise que permite baixos limites de deteção e a fragmentação é facilmente controlada. Os fragmentos dos iões são usados para determinar a estrutura original da molécula [20,24]. Tipicamente nas experiências feitas por *MS/MS*, há a colisão do ião selecionado com o gás inerte como o azoto, hélio e o argon, de onde resultam fragmentos que são analisados com o analisador de massa.

### • Quadrupolo

O primeiro passo é mover os iões carregados da interface de ionização para a área do analisador com elevado vácuo [8,20].

O quadrupolo estabiliza ou destabiliza a trajetória de vôo dos iões com uma mudança de potenciais entre os polos, permitindo um varrimento rápido numa vasta gama de valores de  $m/z$  de aproximadamente até 3000 Da [17]. A corrente contínua destabiliza os iões com maior razão  $m/z$  e provoca um maior desvio na trajetória dos iões, quanto que a corrente alternada destabiliza iões de menor razão  $m/z$ , mas aumentam a amplitude de oscilação.

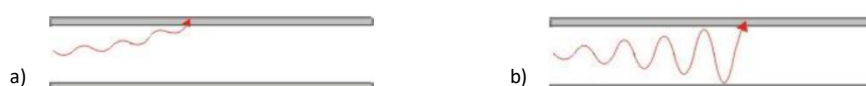


Fig.6 – a) Corrente contínua aplicada quadrupolo; b) Corrente alternada aplicada no quadrupolo

É constituído por quatro hastes cilíndricas (de quartzo) opostas entre si, com a mesma carga aplicada entre elas, enquanto as adjacentes têm cargas opostas aplicadas, fazendo com que os iões sejam bem rearranjados entre si. O espaçamento hiperbólico exato entre as hastes diagonalmente opostas é uma operação crítica no MS.

Entre os tubos é aplicada uma corrente contínua (DC) e corrente alternada (AC) de forma contínua e nas hastes adjacentes com cargas opostas. A intensidade do campo AC/DC é controlada por computador perturbando a onda sinusoidal fazendo passar uma determinada gama de massas de um único fragmento a uma dada frequência [8,17,20,24,26]. Tal facto, permite que os iões percorram um caminho estável até ao detetor e que os fragmentos que não passam numa determinada frequência AC/DC seguem caminhos instáveis acabando por colidir com as paredes do quadrupolo.

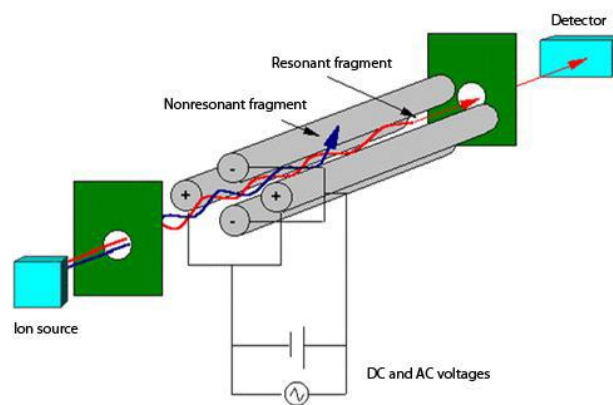


Fig.7 – Demonstração do funcionamento do quadrupolo

O triplo quadrupolo é o analisador de massa mais versátil e combina três quadrupolos, sendo um deles uma célula de colisão.

Durante a reação é possível observar que, no hexapolo os iões moleculares são focados no eixo. No 1º analisador (1º quadrupolo) são separadas as diversas cargas através de um filtro de massa onde os iões de interesse são isolados. Os iões passam por um 2º quadrupolo - célula de colisão, onde ocorre a colisão dos iões anteriormente isolados com um gás de colisão, fazendo com que estes fragmentem.

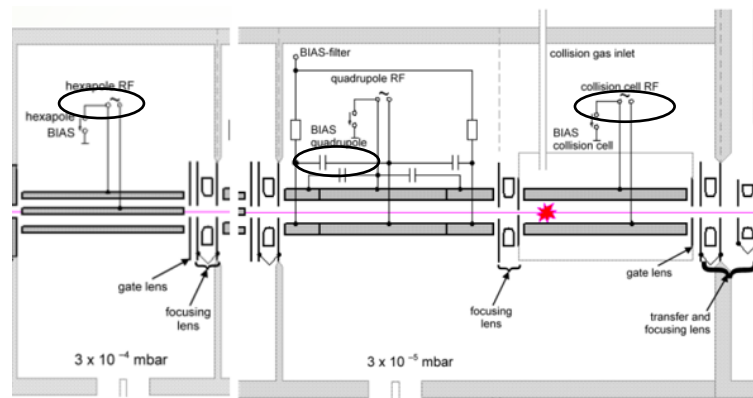


Fig.8 – Analisador de massa: hexapolo, quadrupolo e célula de colisão.

Como já foi referido, o primeiro quadrupolo é usado para selecionar um primeiro ião precursor que é fragmentado na célula de colisão, na qual os iões são acelerados e colidem entre si na presença de um gás de colisão. A energia de colisão com o gás pode variar para permitir diferentes graus de fragmentação [17,26].

O triplo quadrupolo tem vantagens como, o baixo ruído devido à alta sensibilidade e seletividade, boa reprodutibilidade, análise a uma faixa de  $m/z$  até 3000Da capaz de fornecer informações estruturais (seleção, colisão, seleção e deteção).

Os iões estáveis são em seguida expelidos para colidirem com o detetor de iões, produzindo um sinal que é ampliado e enviado para o computador.

- **TOF**

O princípio de operação do TOF-MS envolve a medição do tempo-de-vôo do ião desde o acelerador até ao detetor. A amostra chega ao acelerador ortogonal toda ao mesmo tempo. Aí, é aplicada uma diferença de potencial que faz com que os iões sejam acelerados. Quanto maior for a massa, menor é a velocidade e consequentemente mais tempo demoram a chegar ao detetor [17,20,24,26].

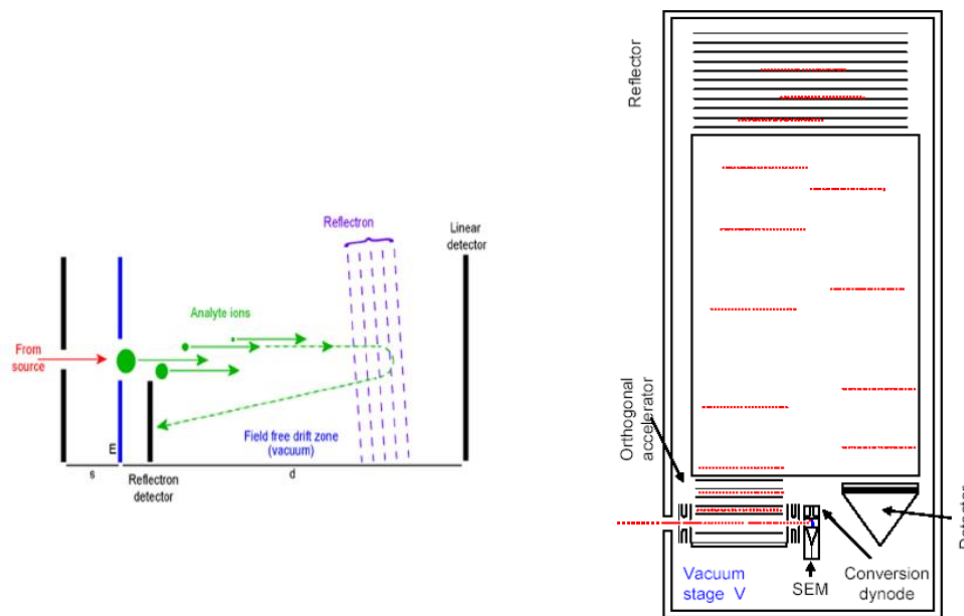


Fig.9 – Princípio de operação do analisador TOF-MS

Se as partículas têm a mesma carga, a energia cinética é bem identificada e as suas velocidades dependem apenas das massas. Quando os iões atravessam a região “campo-livre” entre a fonte, o acelerador e o detetor, são separados devido aos diferentes valores de  $m/z$  [17,26]. Quanto maior for a massa, menor é a velocidade e consequentemente mais tempo demoram a chegar ao detetor.

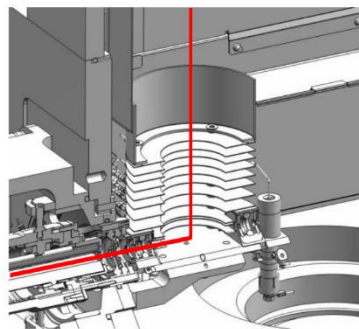


Fig.10 – Princípio de operação do analisador TOF-MS - acelerador ortogonal

No acelerador ortogonal existe a transição de um fluxo de iões contínuo para intermitente. Os iões que chegam ao acelerador são armazenados por algum tempo até que é fechada a passagem para o TOF, e esse conjunto de iões é acelerado [24]. Um analisador TOF com um refletor aumenta a resolução em relação aos lineares, não só porque o percurso aumenta,

mas também por causa da utilização de uma série de campos elétricos que repelem os iões da entrada de volta para a fonte.

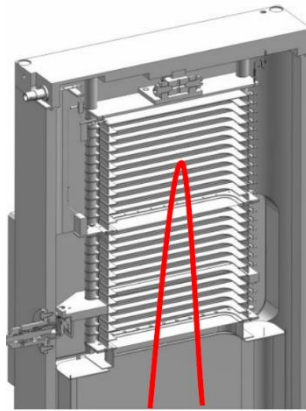


Fig.11 – Princípio de operação do analisador TOF-MS - refletor

A dispersão de energia cinética para os iões de uma dada razão  $m/z$  é corrigida porque os iões com maior EC entram mais fundo na série de placas elétricas, daí o tempo com que voltam a ser refletidos para serem detetados seja maior <sup>[17,20]</sup>.

### 5.3. Detetor

O elemento final do MS é o detetor que regista a corrente produzida quando um ião atinge a superfície do detetor. O sinal produzido no detetor produz um espetro de massa em função da razão  $m/z$ . Normalmente o número de iões que sai do analisador é bastante pequeno e a amplificação é necessária de forma a conseguir obter um sinal razoável <sup>[17,24]</sup>. A sensibilidade, a precisão e o tempo de resposta são parâmetros importantes que distingue os diferentes sistemas de deteção de iões, sendo que estes são normalmente de resposta rápida e de alta precisão.

- **Multiplicador de eletrões**

Quando um ião molecular atinge o detetor, os eletrões recuam e produzem muitos eletrões secundários, ou seja, tem como objetivo a amplificação do sinal elétrico. Tipicamente, o multiplicador de eletrões é o mais usado e baseia-se em dois princípios:

- 1- As partículas que são detetadas têm de ser convertidas em eletrões antes da amplificação sendo normalmente usado um dínodo de conversão.

- 2- Esta amplificação é causada pelos dínodos que aceleram os eletrões a velocidades que lhes permitam gerar mais do que um novo eletrão ao bater no próximo dínodo.

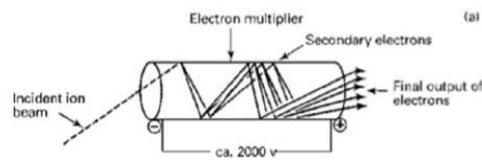


Fig.12 – Princípio de operação do multiplicador de eletrões

A MS usa o princípio de emissão de eletrões secundários para efeitos de amplificação, numa série de dínodos de óxido de alumínio. Os íões a atingir a primeira superfície do dínodo causam uma emissão de eletrões. Estes eletrões são atraídos para o próximo dínodo, de potencial mais elevado, gerando mais eletrões secundários. Os dínodos de alta energia utilizam um campo eletrostático de aceleração para elevar a velocidade dos íões. No feixe de eletrões, estes são acelerados e focados, onde a emissão de eletrões é feita na proporção direta ao número de íões bombeados. A eficiência da MS depende da velocidade da partícula e os eletrões secundários são atraídos ao longo do gradiente elétrico positivo. Cada vez que colidem com a parede, eletrões secundários são expulsos, proporcionando assim a amplificação<sup>[17,24]</sup>.

#### 5.4. Modos de aquisição <sup>[24]</sup>

O espectro de massa resulta num gráfico que apresenta a abundância relativa de cada um dos íões formados em função da sua relação  $m/z$ . Através de um digitalizador ocorre a transformação do sinal analógico (elétrico) num sinal digital. Este sinal produz um espectro com informação relativa aos pesos moleculares de qualquer composto ou fragmentação de íões presentes.

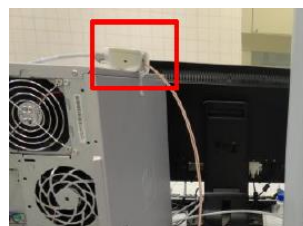


Fig.13 – Digitalizador

O resultado de uma análise por MS é observado por um espectro, onde a abscissa corresponde à razão entre a massa e o número de cargas do ião ( $m/z$ ) e a ordenada está relacionada à sua intensidade. A  $m/z$  tem como unidade o Dalton (Da).

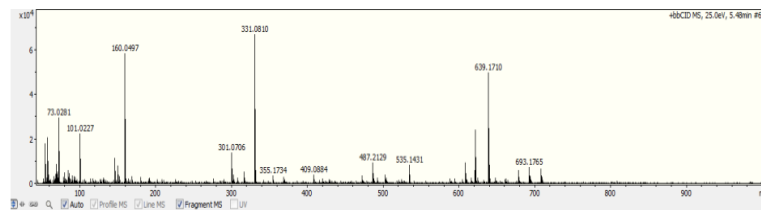


Fig.14 – Exemplificação de um espectro obtido por LC-MS

O espectro de MS também pode, dependendo das condições de aquisição, apresentar sinais referentes à fragmentação do ião originalmente gerado na fonte. Esses fragmentos por sua vez podem ser muito úteis tanto do ponto de vista qualitativo, fornecendo informações auxiliares para elucidação estrutural, quanto quantitativo, no sentido de aumentar a seletividade do método analítico. Finalmente, em virtude do seu caráter universal, o espectro de MS em geral apresenta sinal de outras espécies que são simultaneamente ionizadas ao analito de interesse.

## 6. Análise de componentes principais

### 6.1. Análise multivariada

A química, a partir das experiências que promove, procura extrair informação explicável em termos de teoria que possibilite estabelecer novas teorias, que depois de devidamente validadas pretende, transformar dados em informação e informação em conhecimento <sup>[13]</sup>. Para conseguir esse propósito recorre às ferramentas numéricas disponíveis, dividindo-se estas em dois grupos: um grupo que analisa as variáveis e a informação inerente de forma isolada, estatística univariada, e outro, a análise multivariada que pretende tratar e compreender dados complexos analisando e interpretando o conjunto das variáveis produzidas no seu todo, em simultâneo, com o objetivo de produzir informação <sup>[12,37]</sup>.

A análise multivariada é uma análise exploratória de dados que desempenha um papel crucial no que diz respeito à redução e simplificação de dados sem grande perda de

informação. No caso de ser admissível proceder a uma aproximação em espaços bi- ou tri-dimensionais, é possível uma visualização gráfica aproximada do conjunto de dados [25,39]. O modelo estatístico dos métodos multivariados revela-se, assim, um instrumento poderoso quando se pretende analisar informação proveniente de inúmeras amostras caracterizadas por diversas variáveis [37]. O agrupamento de objetos necessita de uma medida de semelhança, fazendo com que os objetos similares sejam agrupados ou colocados em grupos separados. O PCA é um método exploratório não paramétrico capaz de extrair informação relevante a partir de um conjunto de dados [12].

## 6.2. PCA

A PCA é uma técnica de análise exploratória de dados multivariados desenvolvida por Pearson (1901) e Hotelling (1933) que procura identificar relações das características extraídas de uma matriz de dados. Transforma um conjunto de variáveis correlacionadas num conjunto menor de variáveis independentes, simplificando os dados através da redução do número de variáveis necessárias para os descrever, garantindo a representatividade e importância da base de dados original, através de combinações lineares das variáveis originais, criando novas variáveis designadas por componentes principais, que são facilmente localizadas e analisadas visualmente pois apenas são analisadas as variáveis não correlacionadas [1,25,37].

As componentes principais são ordenadas de modo a que a primeira retenha a maior variação presente nas variáveis originais, a segunda componente menos e assim sucessivamente, sendo que grande parte da variância é retida nas primeiras componentes principais [10]. São ortogonais entre si, e as coordenadas das amostras no novo sistema de referência são designadas por *scores*, enquanto o coeficiente da combinação linear que descreve cada CP, isto é, o impacto das variáveis originais em cada CP, é denominado por *loading*.

$$\text{Var (CP1)} \geq \text{Var (CP2)} \geq \dots \geq \text{Var (CPp)}$$

Fig.15 – Relação entre as variáveis das diferentes componentes principais

Para proceder a análise por PCA é necessário escolher quais as variáveis a considerar, independentemente do agrupamento das amostras, dependendo do objetivo do estudo [11,39]. De seguida escolhe-se a matriz a processar, matriz de covariância – se as variáveis iniciais tiverem a mesma unidade de medida e variâncias próximas – ou a matriz de



correlação – se as variáveis iniciais não tiverem a mesma unidade de medida ou variâncias muito distintas. Por fim, executa-se um programa com o *software* adequado que recorre a várias operações matemáticas para proceder ao tratamento e análise dos dados e escolhe-se o número de componentes principais a analisar, através de um de três critérios distintos ou por conjugação destes: Critério de Pearson, Critério de Kaiser e *Scree Plot*. Para tal, é necessária a que a dimensão da amostra (com variáveis qualitativas) seja “suficientemente grande”, e que exista linearidade das relações entre as variáveis [1, 10, 25, 28, 33, 37].

### • Explicação do processo

Como referido anteriormente, o principal objetivo da PCA é identificar relações entre as características extraídas de uma matriz de dados, reduzindo o seu número de  $p$  variáveis que poderão ser utilizados como indicadores que resumem a informação disponível nas variáveis originais [10,11,25].

A ideia principal é expressar um número  $p$  de variáveis dependentes através de um número  $r$  de variáveis independentes. Considerando uma matriz de dados de  $n$  linhas e  $m$  colunas ( $X_{(n \times m)}$ ), o resultado do PCA é uma rotação do sistema ortogonal de eixos associados às variáveis iniciais (*matriz*  $(n \times m)$ ), em que, após a sua aplicação, se obtém um novo sistema de dados ( $Y_{(n \times 1)}$ ) [33,37].

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mm} \end{bmatrix}$$

Fig.16 – Matriz de dados na análise por PCA

Se existir uma correlação significativa entre as variáveis originais, então o número de componentes principais úteis será mais reduzido do que o número de variáveis originais (geralmente  $r \ll p$ ).

Esta técnica procura encontrar os componentes principais  $CP_1, CP_2, \dots, CP_p$ , que são combinações lineares das variáveis originais,  $X_1, X_2, \dots, X_p$  da matriz  $X$ .

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{pj}X_p$$

Fig.17 – Combinações lineares entre as componentes principais e as variáveis originais da matriz  $X$ , na análise por PCA

Os coeficientes  $a_{11}, a_{12}, \dots, a_{ij}$ , são escolhidos de modo a que as novas variáveis ( $CP_i$ ), ao contrário das variáveis originais ( $X_j$ ), não estejam correlacionadas e que apresentem a máxima variância. Os coeficientes destas combinações lineares são determinados de modo a satisfazerem as condições seguintes [33,37]:

1.  $Var(PC1) \geq Var(PC2) \geq \dots \geq Var(PC_p)$
2. A variância de  $CP_1$  ( $Var(CP_1)$ ) deverá ser tão grande quanto possível, impondo restrições aos coeficientes  $a_{ij}$  tais que:  $a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1$

As variâncias das componentes principais são os valores próprios (*eigenvalues*),  $\lambda$ , da matriz covariância ( $S$ ). O valor próprio associado a cada vetor próprio indica a variância da CP correspondente.

Existem muitas situações em que as variáveis em estudo não são todas medidas na mesma unidade ou são de natureza distinta e é necessário começar por fazer uma normalização dos dados, por forma a ter-se valores médios e/ou variâncias em escalas mais comparáveis. [13,25,28]. As variáveis em estudo passam, assim, a ter todas a mesma variância e a influência das variáveis de variância pequena tende a ser inflacionada enquanto a influência das variáveis de variância elevada tende a ser reduzida. A matriz de covariância do conjunto destas novas variáveis é igual à matriz de correlação do conjunto de variáveis iniciais, dado que [1,33,37]:

$$Cov\left(\frac{X_i}{\sigma_i}, \frac{X_j}{\sigma_j}\right) = \frac{Cov(X_i, X_j)}{\sigma_i \sigma_j} = Corr(X_i, X_j)$$

Fig.18 – Matriz covariância das novas variáveis na análise por PCA

A redução de dimensionalidade consegue-se considerando as componentes principais de maior variância. Ao ordenarmos as componentes principais por ordem decrescente da sua variância podemos tomar como a componente principal de maior importância aquela que tiver associada a maior variância, a mais representativa dos dados originais [10,25,28,39].

Existem vários critérios que podem ser usados para a escolha do número de componentes principais: Critério de *Pearson* (ou regra dos 80%), Critério de *Kaiser* ( $\lambda > 1$ ) e *Scree plot*.

O Critério de *Pearson* (ou regra dos 80%) pode ser utilizado quando são utilizadas as matrizes de covariância ou de correlação. O número de componentes principais é escolhido

até recuperarmos mais de 80% da informação total ou da variabilidade total de forma a reter as primeiras componentes principais de modo a que [33,37]:

$$\frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.80$$

Fig.19 – Demonstração do critério de Pearson

O critério de *Kaiser-Guttman* é utilizado com a matriz de correlação (o critério anteriormente descrito também é uma possibilidade), em que ( $p$ ) é o número de variáveis e ( $r$ ) o número de componentes. Neste caso, devem ser consideradas apenas as componentes com valor próprio superior a 1 (o valor unitário é a média do conjunto de valores próprios) [25,33,37].

$$\bar{\lambda} = \frac{1}{p} \sum_{j=1}^p \lambda_j$$

Fig.20 – Demonstração do critério de Kaiser-Guttman

O critério de *Scree Plot* ou teste de *Cattell* (Cattell, 1966), que é um método que consiste na observação do gráfico dos *eigenvalues*, gráfico onde os pontos de maior declive são indicativos do número total de componentes. Por meio da análise do gráfico, é possível observar quais fatores apresentam maiores *eigenvalues*, sendo, portanto, responsáveis por uma maior variância explicada [13,33,37].

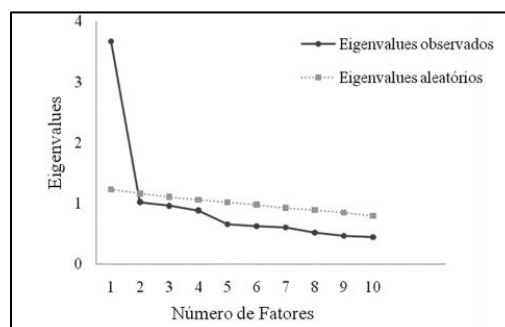


Fig.21 – Exemplo de um gráfico obtido pelo critério de *Scree Plot* ou teste de *Cattell*

No PCA, uma matriz é decomposta em uma soma de produtos externos associados a um autovalor ( $\lambda_r$ ) que indica a importância dos vetores associados na representação da base de dados original. Essa relevância de informação contida no autovalor, faz com que se torne um critério de seleção do número de componentes, tendo em vista ter uma variância explicada pelo modelo de acordo com a quantidade de autovalores <sup>[1,25,33,39]</sup>.

- **Vantagens e desvantagens do método PCA**

Métodos de primeira ordem como PCA empregam dados vetoriais com a vantagem de produzir robustez à análise, identificar a presença de interferentes ou amostras com erros elevados e ainda, os desvios de um processo. Nesses casos, a quantificação é possível, mesmo na presença de interferentes, desde que esses sejam previamente levados em consideração na construção do modelo de calibração. No entanto, para a boa obtenção deste método, é necessário um elevado número de amostras para que os possíveis interferentes estejam identificados durante a calibração <sup>[11,39]</sup>. Esse pressuposto mostra-se bastante atraente e proporciona uma série de vantagens, como o melhor entendimento das possibilidades de um sistema, a partir do tipo de resposta obtida pelo uso de um ou outro instrumento ou metodologia analítica. Outra vantagem é a otimização dos métodos existentes, conduzindo a pesquisa científica o desenvolvimento de ferramentas mais poderosas. A escolha primordial neste tipo de modelos é a escolha do número ideal de fatores para a decomposição <sup>[1,33]</sup>.

i) é um critério objetivo, no sentido de não estar sujeito a apreciações diferenciadas por parte de utilizadores diferentes;

ii) é de cálculo muito fácil;

iii) conta com todos os valores próprios através da sua variância, utilizando a totalidade da informação disponível;

iv) tem uma justificação geométrica, independentemente da natureza dos dados em estudo;

v) não exige qualquer hipótese distribucional.

Algumas destas vantagens (em particular as últimas) poderão ser encaradas como desvantagens em situações particulares, como quando a natureza da aplicação aconselha outros critérios de escolha. Mas na generalidade das situações onde uma análise fatorial se aplica como técnica exploratória, a inexistência de hipóteses prévias sobre a natureza dos dados é, sem dúvida, uma vantagem <sup>[12,33,37]</sup>.

Concluindo, na redução da dimensionalidade de grandes matrizes de dados as  $p$  variáveis originais são substituídas por um outro subconjunto de  $r$  variáveis não correlacionadas, de menor dimensão, designadas por componentes principais, com a menor perda de informação possível [10,25,28]. Acresce que, as novas variáveis, as componentes principais, não são correlacionadas o que simplifica a análise pois em vez de se analisar um número elevado de variáveis com uma estrutura inter-relacional complexa, as originais, analisam-se somente algumas variáveis não correlacionadas. A aplicação deste método permite, assim, a simplificação e redução da dimensão original dos dados, a modelação, a deteção de *outliers*, a seleção de variáveis importantes num determinado sistema, a classificação e a previsão [12,33,37].

## **Capítulo II**

### **DESCRIÇÃO DO TRABALHO**

## 1. Apresentação e enquadramento do projeto

O vinho é essencial para a vida e cultura europeia, sendo a União Europeia a maior produtora de vinhos do mundo e líder mundial na exportação de produtos vitivinícolas. A sustentabilidade da produção de vinho é definida pela Organização Internacional da Vinha e do Vinho (OIV) como a estratégia global que inclui todas as etapas do ciclo produtivo do vinho. Em Portugal, o vinho tem vindo a adquirir uma grande projeção nos panoramas nacional e internacional, acumulando vários prémios e distinções.

Ao longo dos tempos, os apreciadores de vinho raramente beberam um produto puro feito exclusivamente de uvas. A natureza nem sempre oferecia aos produtores e consumidores um produto com as qualidades que se foram exigindo ao longo dos tempos. A solução através de conhecimentos práticos e empíricos era adicionar produtos que mantivessem ou melhorassem as suas qualidades. A química do vinho é bastante complexa e ainda hoje não é completamente conhecida, embora as técnicas espectroscópicas tenham permitido enormes avanços na compreensão dos processos que ocorrem no vinho.

É através da análise química que é possível detetar a maior parte das fraudes no fabrico do vinho. A evolução das fraudes, seja no aspeto qualitativo ou no aspeto quantitativo, tem atingido mundialmente as empresas produtoras de vinho, ocasionando perdas substanciais nos respetivos patrimónios. Como tal, é necessário investigar mais sobre esta temática que tanto preocupa os produtores e os consumidores.

Assim, este projeto que decorreu na Sogrape Vinhos S.A, teve como objetivo a criação de uma base de dados de vinhos por região vitivinícola, em que se pretendeu identificar a proveniência dos vinhos comprados, detetar fraudes e contaminações acidentais. As principais tarefas a desenvolver durante o estágio foram desenhar o protocolo para a criação da base de dados e testá-lo, sendo estas as maiores necessidades atuais da empresa.

## 2. Apresentação e caracterização da instituição acolhedora de estágio <sup>[35]</sup>

Foi através da visão, caráter e personalidade de Fernando van Zeller Guedes que em 1942 foi fundada a Sogrape Vinhos S.A., uma empresa que se dedica ao cultivo, produção e exportação de vinho. O grande objetivo da empresa era ser diferente, capaz de divulgar e impor os vinhos portugueses nos mercados internacionais e foi então que resultou a criação da primeira marca portuguesa de vinhos global – Mateus Rosé – que devido ao seu enorme sucesso serviu de impulso para o crescimento e afirmação da empresa.

Fernando Guedes sabia que Inglaterra era a “montra do mundo”, mas até à conquista apostou no mercado de colónias portuguesas em África, principalmente Angola e Moçambique. Após a popularidade alcançada no Reino Unido é que a Sogrape adquiriu os meios financeiros e o conhecimento dos mercados para a empresa alargar a sua presença em outras regiões, como foi o caso da Região Demarcada do Dão, Douro, Bairrada e Vinhos Verdes. Apostaram fortemente nos setores dos Vinhos do Porto e Douro com a aquisição de AA. Ferreira S.A., mais tarde da Offley e por último adquiriu a totalidade dos ativos da Sandeman. Foram ainda feitos grandes investimentos na região do Alentejo, após compra da Herdade do Peso.

Hoje, liderada pela 3ª geração da família fundadora, a Sogrape soube crescer de forma sustentada e com um permanente investimento na valorização do seu património e capital humano, sendo hoje detentora de mais de 830 hectares de área de vinha nas principais regiões vinícolas portuguesas, 18 quintas, 15 centros de vinificação e 9 linhas de engarrafamento.

Atualmente o processo produtivo da Sogrape é bastante valioso, apostando em vinhos de elevada qualidade com diversos perfis e sabores. Para Fernando Guedes, atual presidente da Sogrape Vinhos, “este reconhecimento é motivo de grande orgulho e satisfação, refletindo o empenho e a paixão que colocamos diariamente na criação de vinhos de qualidade que correspondem às exigências dos distintos consumidores”.

### **3. Organização da tese**

Na introdução teórica é apresentada uma breve caracterização das diferentes regiões vitivinícolas portuguesas e é feita uma pequena referência a alguns casos de fraudes em vinhos. São descritos os princípios do método de análise – LC-MS – e do método do tratamento dos dados – PCA. É apresentado o enquadramento do trabalho e os motivos pelos quais é importante desenvolver bases de dados, para ser possível identificar fraudes e outros tipos de contaminações no vinho.

De seguida, são apresentados os reagentes e equipamento utilizados, bem como o procedimento adotado na preparação da amostra assim como as condições utilizadas na análise por LC-MS.

Na secção correspondente aos resultados e discussão são apresentados espectros originais obtidos bem como a otimização das condições no tratamento dos dados por PCA, recorrendo ao programa ProfileAnalysis e os resultados por este obtidos. Seguidamente, às conclusões obtidas é feita uma avaliação do trabalho realizado, principalmente no que diz respeito aos objetivos cumpridos e às principais limitações.



## **Capítulo III**

### **MATERIAIS E MÉTODOS**

Neste capítulo, são apresentados os reagentes e equipamento utilizados, bem como o procedimento adotado na preparação da amostra assim como as condições utilizadas na análise por LC-MS.

## 1. Amostras e processo de amostragem

O protocolo de amostragem tinha que ser baseado no mínimo em 100 amostras representativas de 3 anos de colheita (2013, 2014 e 2015), por região e tipo de vinho. Mais tarde, iria proceder-se á simulação de pelo menos 3 casos de adulteração.

O presente estudo foca principalmente as regiões do vinho do Porto, Douro, Dão e Alentejo, sendo que também foram analisadas algumas amostras da região dos vinhos Verdes, num total de 414 amostras.

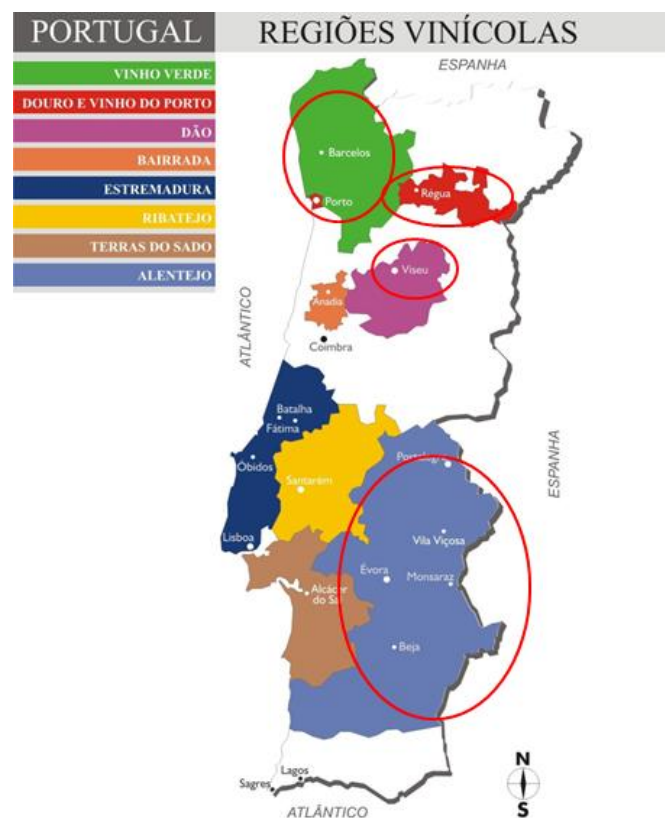


Fig.22 – Mapa das regiões vinícolas portuguesas em estudo (assinaladas no respetivo mapa)

Tabela 1: Número de amostras analisadas no ano de 2013 nas diferentes regiões em estudo

Região	Ano	Nº de amostras
Douro	2013	149
Dão		37
Porto		5
Alentejo		11
Verdes		0
Total		202

Tabela 2: Número de amostras analisadas no ano de 2014 nas diferentes regiões em estudo

Região	Ano	Nº de amostras
Douro	2014	8
Dão		2
Porto		45
Alentejo		4
Verdes		0
Total		59

Tabela 3: Número de amostras analisadas no ano de 2015 nas diferentes regiões em estudo

Região	Ano	Nº de amostras
Douro	2015	24
Dão		19
Porto		49
Alentejo		35
Verdes		25
Total		153

Todas as amostras que chegaram ao laboratório foram recolhidas sob a orientação do enólogo responsável pela região em estudo, de forma a garantir uma amostragem bem diversificada e distribuída.

## 2. Reagentes

Os solventes usados foram o propan-2-ol (*Fluka, LC-MS CHROMASOLV* com um grau de pureza superior a 99,9%), o acetonitrilo (*hypergrade for liquid chromatography (LC-MS), Merck, 99.9%*), o ácido fórmico (*Merck, com um grau de pureza entre 98-100%*), e o hidróxido de sódio (*VWR® - Chemicals Prolabo, 99%*).

A água utilizada era ultrapura, tratada num sistema de purificação de água *Milli-Q Advantage 10*. Resistividade de 18MΩ e com TOC <5 ppb.

Os padrões internos foram o trifenilfosfato e o ciclamato de sódio.

## 3. Equipamento

O LC-MS é um *Dionex Ultimate 3000 RSLC / maXis impact*, com um *autosampler Dionex UltiMate 3000 – WPS-3000RS*.

O MS é equipado com uma fonte de ionização por *eletrospray Apolo* e o analisador de massa é um Q-TOF (triplo quadrupolo). A coluna utilizada foi a *Acclaim RSLC 120* de fase reversa C18 (100x2,1mm, 1,7μm).

A aquisição dos dados foi feita recorrendo aos programas *Chromeleon*, *otofControl* e *HyStar*.

## 4. Preparação da amostra

Todas as amostras chegam ao laboratório em pequenos frascos e estes são identificados de acordo com o seu conteúdo, tipo de vinho, o ano e a região de que são provenientes.



Fig.23 – Chegada das amostras ao laboratório

Todos os dados das amostras, já referidos anteriormente, são devidamente registados num ficheiro do Excel.

Tabela 4: Exemplo da identificação de uma amostra após a sua chegada ao laboratório

Amostra nº	Tipo	Ano	Região	Variedade das uvas	Código	Data de receção	Data de análise
1	Red	2013	Douro	TF (57%); TC (22%); DIV (21%)	VTD13_1	4-2-16	5-2-16

Para melhorar a sua identificação foi ainda feita uma matriz de códigos, como é apresentada na tabela 5.

Tabela 5: Matriz de códigos usada para uma melhor identificação das diversas amostras

Matriz de códigos										
Tipo de produto	Região	Ano	Código	Ano	Código	Ano	Código			
Vinho Tinto	Alentejo	2013	VTA13_X	2014	VTA14_X	2015	VTA15_X			
	Dão	2013	VTDA13_X	2014	VTDA14_X	2015	VTDA15_X			
	Douro	2013	VTD13_X	2014	VTD14_X	2015	VTD15_X			
	Porto	2013	VTP13_X	2014	VTP14_X	2015	VTP15_X			
Vinho Branco	Alentejo	2013	VBA13_X	2014	VBA14_X	2015	VBA15_X			
	Dão	2013	VBDA13_X	2014	VBDA14_X	2015	VBDA15_X			
	Douro	2013	VBD13_X	2014	VBD14_X	2015	VBD15_X			
	Verde	2013	VBV13_X	2014	VBV14_X	2015	VBV15_X			
	Porto	2013	VBP13_X	2014	VBP14_X	2015	VBP15_X			
Vinho Rosé	Alentejo	2013	VRA13_X	2014	VRA14_X	2015	VRA15_X			
	Dão	2013	VRDA13_X	2014	VRDA14_X	2015	VRDA15_X			
	Douro	2013	VRD13_X	2014	VRD14_X	2015	VRD15_X			
	Porto	2013	VRP13_X	2014	VRP14_X	2015	VRP15_X			

Após a sua identificação, as amostras são centrifugadas durante cinco minutos e filtradas através de uma seringa com filtros de polipropileno de 0,2µm. São colocadas em frascos selados para posteriormente se proceder à sua injeção no LC-MS.



Fig.24 – Selagem dos frascos antes da análise por LC-MS

## 5. Análise por LC-MS

A análise das diferentes amostras de vinho para este projeto foi concebida de forma a obter o máximo de informação a partir da amostra, a fim de permitir o reprocessamento sem outras medidas de aquisição de dados.

Cada amostra foi analisada em duplicado, em modo de ionização positivo e negativo.

A aquisição dos dados foi feita recorrendo aos programas *Chromeleon*, *otofControl* e *HyStar*.

### • Método UHPLC

A análise por UHPLC foi realizada a uma temperatura de 30°C e foi feita por injeção direta (3µL) numa coluna *Acclaim RSLC 120* de fase reversa C18 (100x2,1mm 1,7µm). Os solventes utilizados foram A, água/ácido fórmico e B, acetonitrilo/ácido fórmico.

Tabela 6: Descrição dos constituintes dos solventes A e B, utilizados na fase móvel do HPLC

Solvente	Descrição
A	• Água (Milli-Q 18MΩ TOC < 5 ppb)
	• Ácido fórmico 0,1 % (v/v)
B	• Acetonitrilo
	• Ácido fórmico 0,1 % (v/v)

Foram utilizadas as duas fases móveis descritas anteriormente, num fluxo de 0,3mL/min e operou-se em modo gradiente, sendo as condições apresentadas na tabela 7. O solvente nunca deve ser 100% aquoso, devido á supressão de fase.

Tabela 7: Percentagem de cada um dos solventes usados na fase móvel ao longo do tempo de aquisição com um fluxo de 0,3mL/min

Tempo (min)	Fluxo (mL/min)	% A	% B	Observações
0	0,3	98	2	
1	0,3	98	2	
2	0,3	75	25	Rampa
10	0,3	0	100	
12,5	0,3	0	100	Limpeza
12,5	0,3	98	2	
15	0,3	98	2	Equilibration

#### • Método MS

A condições operatórias usadas já tinham sido testadas anteriormente num outro projeto da empresa “OxiPorto”, cujo objetivo era identificar e conseguir separar as amostras pelos diferentes anos, dos diversos vinhos do Porto em estudo.

Como tal, as condições já estavam otimizadas visto que o equipamento a usar e o objetivo da análise era o mesmo. Foram utilizadas as duas fases móveis (A e B) já descritas anteriormente. As outras principais condições de análise estão descritas na tabela 8.

Tabela 8: Principais condições de análise no método MS

Parâmetro	Valor
Ionização	Positivo / negativo
Intervalo de massa	50-1250 <i>m/z</i>
Voltagem capilar	2500 V/3000 V
Pressão do nebulizador	2,0 bar
Temperatura do gás	200°C
Gás nebulizador	N <sub>2</sub>
Modo de calibração	Quadrático + HPC
Calibrante	Formato de sódio
Tempo de corrida	15 minutos

Temperatura da amostra	5°C [4-45]°C
Temperatura da coluna	30°C [25-35]°C
Limites de pressão	0-800 bar
Energia do ião no quadrupolo	8.0 eV
Energia na célula de colisão	10.0 eV
<i>End Plate</i>	500V

## • Calibrações

### Calibração do equipamento

Para a calibração do equipamento é feito um *auto tune* mensalmente de forma a ajustar as voltagens que são aplicadas nas diferentes etapas do processo de análise, pois o detetor degrada-se com o tempo/uso. Para manter a sensibilidade e o sinal constante ao longo do tempo é necessário efetuar ajustes na voltagem do detetor. É comprada uma solução designada “*Tuning Mix*” que contém uma série de compostos conhecidos. É ainda feita uma calibração das massas diariamente com formato de sódio.

### Calibração das amostras

A cada amostra é adicionado o calibrante interno com um *loop* de 20µL através da válvula de 6 portas. Na posição *waste* é feita uma ligação direta do UHPLC ao MS e a bomba de seringa enche o *loop* de calibrante. Na posição *source*, o calibrante do *loop* é arrastado pela corrente do HPLC para o MS e a bomba de seringa está desligada do *loop*. A calibração da amostra é feita também com o formato de sódio e com a *lock mass*.



Fig.25 – Calibração da amostra através da válvula de 6 portas



## Padrão interno

A fonte de ionização quando opera em modo positivo, o padrão interno usado é o trifenilfosfato e em modo negativo é o ciclamato de sódio.

Tabela 9: Propriedades químicas dos calibrantes e dos padrões internos usados em LC-MS

Calibrante	Nome científico	Fórmula química	Massa molar (g mol <sup>-1</sup> )
<i>“Lock mass calibrant solution”</i>	Hexakis(2,2-diethoxy)phosphazene	C <sub>12</sub> H <sub>18</sub> F <sub>12</sub> N <sub>3</sub> O <sub>6</sub> P <sub>3</sub>	621,19
*Formato de sódio	Formato de sódio	HCOONa	68,01
Padrão interno (modo)	Nome científico	Fórmula química	Massa molar g mol <sup>-1</sup>
Positivo	Trifenilfosfato (fosfato de trifenila)	C <sub>18</sub> H <sub>15</sub> O <sub>4</sub> P	326,28
Negativo	Ciclamato de sódio (N'-ciclo-hexil-sulfamato de sódio)	NaC <sub>6</sub> H <sub>12</sub> SNO <sub>3</sub>	201,22

\*250mL H<sub>2</sub>O + 250mL isopropanol + 250μL ácido fórmico + 0,5mL de NaOH 1N

## **Capítulo IV**

### **RESULTADOS E DISCUSSÃO**

## 1. Calibração dos espectros de LC-MS

O software utilizado no processamento dos dados foi o *DataAnalysis*. Permite uma visualização e processamento (calibração) dos cromatogramas e dos espectros de massa obtidos. As condições de calibração usadas quer em modo de ionização positivo, quer em modo de ionização negativo são apresentadas na tabela 10.

Tabela 10: Condições usadas no programa *DataAnalysis* para a calibração dos espectros

Parâmetro	Valor
Lista de massas de referência	Formato de sódio
Intervalo de pesquisa	0,5 m/z
Modo de calibração	HPC
Calibrante	Formato de sódio
<i>Lock mass (pos/neg)</i>	622/556
Intensidade mínima do sinal do calibrante	5000

Após a calibração dos espectros, e sobrepondo os resultados dos diferentes vinhos das diversas regiões, observa-se o cromatograma do pico base (*BPC- Base peak chromatogram*) sendo que cada ponto do cromatograma (min) corresponde ao ponto de maior intensidade (m/z). Na figura 26, as diferentes cores são representativas de vinhos de cada região sendo que há picos que não são comuns, o que são indicativos de compostos diferenciadores. Cada ponto do cromatograma corresponde a um espectro de massa diferente.

É ainda possível visualizar o *total ion chromatogram* (TIC), que a cada ponto do cromatograma (min) corresponde à soma de todas as intensidades (m/z) e o *Extracted ion chromatogram* (EIC), em que cada ponto apresenta a intensidade da massa escolhida.

A figura 26 representa a sobreposição de quatro amostras diferentes de 2013 quando o LC-MS operou em modo positivo.

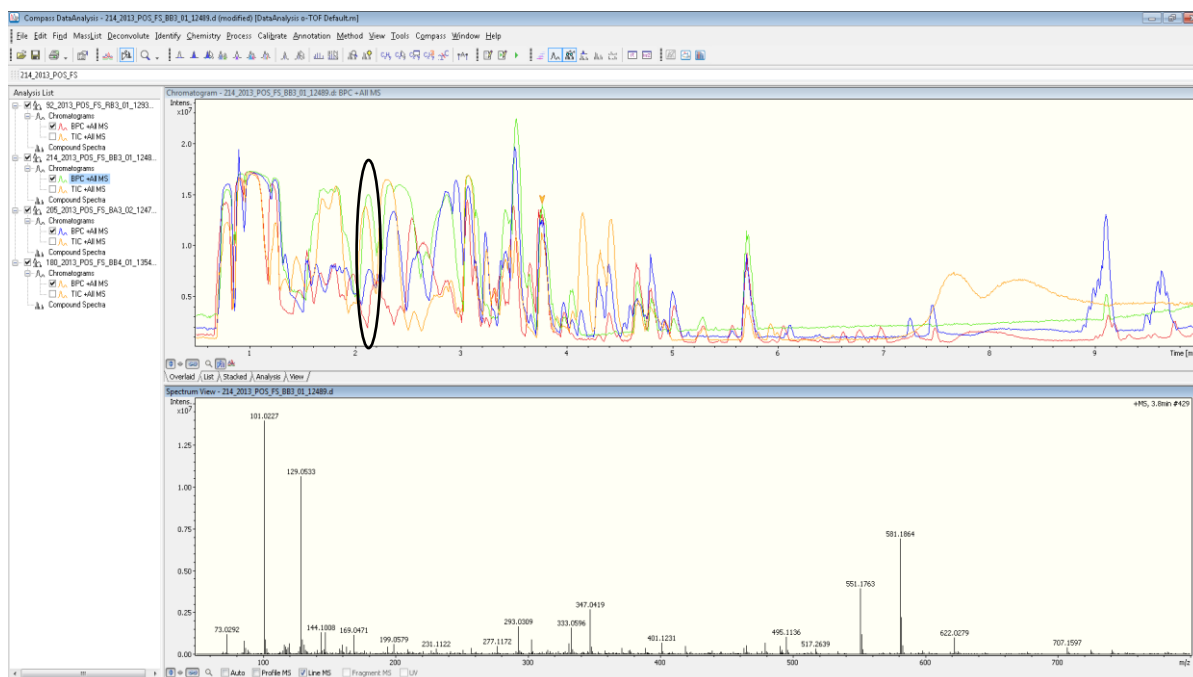


Fig.26 – Sobreposição de quatro cromatogramas (BPC) ao fim de calibrados, modo positivo.

Na sobreposição dos cromatogramas, a linha a vermelho é representativa da amostra do Douro, a linha a amarelo da amostra do Dão, a linha a azul corresponde a um vinho do Porto e a linha a verde a uma amostra oriunda do Alentejo. O espetro de massa observado corresponde ao composto presente na amostra 214 do Alentejo, perto do minuto 2,1 (assinalado na figura 26).

Ao fim de verificar as diferenças existentes nas regiões, o objetivo é pesquisar quais os compostos diferenciadores, em todas as amostras.

## 2. PCA

O software utilizado no processamento estatístico foi o *ProfileAnalysis*. Os parâmetros usados estão descritos na tabela 11.

Tabela 11: Condições usadas no programa *ProfileAnalysis*

Parâmetro	Valor
Tempo de aquisição	0.5 – 10 min
Intervalo de massa	50 – 1250 m/z
<i>Bucketing</i>	<i>Advanced Bucketing</i>

• Tempo	0.1 / 0.2min (neg/pos)
• Massa	3 / 50mDa (neg/pos)
• Separação dos <i>buckets</i> em vários compostos	Ligado
Normalização	Não
Filtro do <i>bucket</i>	Ligado
• Value count of <i>bucket</i>	≥10
• Substituição dos valores em falta	Valor mínimo do <i>bucket</i>
Transformação do <i>bucket</i>	Não

## • Amostras de 2013, quando ionizadas em modo negativo

Após a otimização de todas as condições acima mencionadas é obtida a *bucket table* que relaciona o tempo com a razão *m/z* (figura 27) em que os valores assinalados a vermelhos são referentes aos *outliers*. Com a exclusão dos *outliers* do modelo (PCA), não são consideradas as diferenças numa análise que possivelmente possa ter corrido mal, mas sim os compostos que diferem de amostra para amostra.

Bucket Table 1		Bucket Table		153_2013_N		153_2013_N		154_2013_N		154_2013_N		155_2013_N		155_2013_N		156_2013_N		156_2013_N		157_2013_N		157_2013_N		158_2013_N		158_2013_N		159_2013_N		159_2013_N		160_2013_N		160_2013_N		161_2013_N		161_2013_N		162_2013_N		162_2013_N		163_2013_N		163_2013_N		164_2013_N		164_2013_N		165_2013_N		165_2013_N		166_2013_N		166_2013_N		167_2013_N		167_2013_N		168_2013_N		168_2013_N		169_2013_N		169_2013_N		170_2013_N		170_2013_N		171_2013_N		171_2013_N		172_2013_N		172_2013_N		173_2013_N		173_2013_N		174_2013_N		174_2013_N		175_2013_N		175_2013_N		176_2013_N		176_2013_N		177_2013_N		177_2013_N		178_2013_N		178_2013_N		179_2013_N		179_2013_N		180_2013_N		180_2013_N		181_2013_N		181_2013_N		182_2013_N		182_2013_N		183_2013_N		183_2013_N		184_2013_N		184_2013_N		185_2013_N		185_2013_N		186_2013_N		186_2013_N		187_2013_N		187_2013_N		188_2013_N		188_2013_N		189_2013_N		189_2013_N		190_2013_N		190_2013_N		191_2013_N		191_2013_N		192_2013_N		192_2013_N		193_2013_N		193_2013_N		194_2013_N		194_2013_N		195_2013_N		195_2013_N		196_2013_N		196_2013_N		197_2013_N		197_2013_N		198_2013_N		198_2013_N		199_2013_N		199_2013_N		200_2013_N		200_2013_N		201_2013_N		201_2013_N		202_2013_N		202_2013_N		203_2013_N		203_2013_N		204_2013_N		204_2013_N		205_2013_N		205_2013_N		206_2013_N		206_2013_N		207_2013_N		207_2013_N		208_2013_N		208_2013_N		209_2013_N		209_2013_N		210_2013_N		210_2013_N		211_2013_N		211_2013_N		212_2013_N		212_2013_N		213_2013_N		213_2013_N		214_2013_N		214_2013_N		215_2013_N		215_2013_N		216_2013_N		216_2013_N		217_2013_N		217_2013_N		218_2013_N		218_2013_N		219_2013_N		219_2013_N		220_2013_N		220_2013_N		221_2013_N		221_2013_N		222_2013_N		222_2013_N		223_2013_N		223_2013_N		224_2013_N		224_2013_N		225_2013_N		225_2013_N		226_2013_N		226_2013_N		227_2013_N		227_2013_N		228_2013_N		228_2013_N		229_2013_N		229_2013_N		230_2013_N		230_2013_N		231_2013_N		231_2013_N		232_2013_N		232_2013_N		233_2013_N		233_2013_N		234_2013_N		234_2013_N		235_2013_N		235_2013_N		236_2013_N		236_2013_N		237_2013_N		237_2013_N		238_2013_N		238_2013_N		239_2013_N		239_2013_N		240_2013_N		240_2013_N		241_2013_N		241_2013_N		242_2013_N		242_2013_N		243_2013_N		243_2013_N		244_2013_N		244_2013_N		245_2013_N		245_2013_N		246_2013_N		246_2013_N		247_2013_N		247_2013_N		248_2013_N		248_2013_N		249_2013_N		249_2013_N		250_2013_N		250_2013_N		251_2013_N		251_2013_N		252_2013_N		252_2013_N		253_2013_N		253_2013_N		254_2013_N		254_2013_N		255_2013_N		255_2013_N		256_2013_N		256_2013_N		257_2013_N		257_2013_N		258_2013_N		258_2013_N		259_2013_N		259_2013_N		260_2013_N		260_2013_N		261_2013_N		261_2013_N		262_2013_N		262_2013_N		263_2013_N		263_2013_N		264_2013_N		264_2013_N		265_2013_N		265_2013_N		266_2013_N		266_2013_N		267_2013_N		267_2013_N		268_2013_N		268_2013_N		269_2013_N		269_2013_N		270_2013_N		270_2013_N		271_2013_N		271_2013_N		272_2013_N		272_2013_N		273_2013_N		273_2013_N		274_2013_N		274_2013_N		275_2013_N		275_2013_N		276_2013_N		276_2013_N		277_2013_N		277_2013_N		278_2013_N		278_2013_N		279_2013_N		279_2013_N		280_2013_N		280_2013_N		281_2013_N		281_2013_N		282_2013_N		282_2013_N		283_2013_N		283_2013_N		284_2013_N		284_2013_N		285_2013_N		285_2013_N		286_2013_N		286_2013_N		287_2013_N		287_2013_N		288_2013_N		288_2013_N		289_2013_N		289_2013_N		290_2013_N		290_2013_N		291_2013_N		291_2013_N		292_2013_N		292_2013_N		293_2013_N		293_2013_N		294_2013_N		294_2013_N		295_2013_N		295_2013_N		296_2013_N		296_2013_N		297_2013_N		297_2013_N		298_2013_N		298_2013_N		299_2013_N		299_2013_N		300_2013_N		300_2013_N		301_2013_N		301_2013_N		302_2013_N		302_2013_N		303_2013_N		303_2013_N		304_2013_N		304_2013_N		305_2013_N		305_2013_N		306_2013_N		306_2013_N		307_2013_N		307_2013_N		308_2013_N		308_2013_N		309_2013_N		309_2013_N		310_2013_N		310_2013_N		311_2013_N		311_2013_N		312_2013_N		312_2013_N		313_2013_N		313_2013_N		314_2013_N		314_2013_N		315_2013_N		315_2013_N		316_2013_N		316_2013_N		317_2013_N		317_2013_N		318_2013_N		318_2013_N		319_2013_N		319_2013_N		320_2013_N		320_2013_N		321_2013_N		321_2013_N		322_2013_N		322_2013_N		323_2013_N		323_2013_N		324_2013_N		324_2013_N		325_2013_N		325_2013_N		326_2013_N		326_2013_N		327_2013_N		327_2013_N		328_2013_N		328_2013_N		329_2013_N		329_2013_N		330_2013_N		330_2013_N		331_2013_N		331_2013_N		332_2013_N		332_2013_N		333_2013_N		333_2013_N		334_2013_N		334_2013_N		335_2013_N		335_2013_N		336_2013_N		336_2013_N		337_2013_N		337_2013_N		338_2013_N		338_2013_N		339_2013_N		339_2013_N		340_2013_N		340_2013_N		341_2013_N		341_2013_N		342_2013_N		342_2013_N		343_2013_N		343_2013_N		344_2013_N		344_2013_N		345_2013_N		345_2013_N		346_2013_N		346_2013_N		347_2013_N		347_2013_N		348_2013_N		348_2013_N		349_2013_N		349_2013_N		350_2013_N		350_2013_N		351_2013_N		351_2013_N		352_2013_N		352_2013_N		353_2013_N		353_2013_N		354_2013_N		354_2013_N		355_2013_N		355_2013_N		356_2013_N		356_2013_N		357_2013_N		357_2013_N		358_2013_N		358_2013_N		359_2013_N		359_2013_N		360_2013_N		360_2013_N		361_2013_N		361_2013_N		362_2013_N		362_2013_N		363_2013_N		363_2013_N		364_2013_N		364_2013_N		365_2013_N		365_2013_N		366_2013_N		366_2013_N		367_2013_N		367_2013_N		368_2013_N		368_2013_N		369_2013_N		369_2013_N		370_2013_N		370_2013_N		371_2013_N		371_2013_N		372_2013_N		372_2013_N		373_2013_N		373_2013_N		374_2013_N		374_2013_N		375_2013_N		375_2013_N		376_2013_N		376_2013_N		377_2013_N		377_2013_N		378_2013_N		378_2013_N		379_2013_N		379_2013_N		380_2013_N		380_2013_N		381_2013_N		381_2013_N		382_2013_N		382_2013_N		383_2013_N		383_2013_N		384_2013_N		384_2013_N		385_2013_N		385_2013_N		386_2013_N		386_2013_N		387_2013_N		387_2013_N		388_2013_N		388_2013_N		389_2013_N		389_2013_N		390_2013_N		390_2013_N		391_2013_N		391_2013_N		392_2013_N		392_2013_N		393_2013_N		393_2013_N		394_2013_N		394_2013_N		395_2013_N		395_2013_N		396_2013_N		396_2013_N		397_2013_N		397_2013_N		398_2013_N		398_2013_N		399_2013_N		399_2013_N		400_2013_N		400_2013_N		401_2013_N		401_2013_N		402_2013_N		402_2013_N		403_2013_N		403_2013_N		404_2013_N		404_2013_N		405_2013_N		405_2013_N		406_2013_N		406_2013_N		407_2013_N		407_2013_N		408_2013_N		408_2013_N		409_2013_N		409_2013_N		410_2013_N		410_2013_N		411_2013_N		411_2013_N		412_2013_N		412_2013_N		413_2013_N		413_2013_N		414_2013_N		414_2013_N		415_2013_N		415_2013_N		416_2013_N		416_2013_N		417_2013_N		417_2013_N		418_2013_N		418_2013_N		419_2013_N		419_2013_N		420_2013_N		420_2013_N		421_2013_N		421_2013_N		422_2013_N		422_2013_N		423_2013_N		423_2013_N		424_2013_N		424_2013_N		425_2013_N		425_2013_N		426_2013_N		426_2013_N		427_2013_N		427_2013_N		428_2013_N		428_2013_N		429_2013_N		429_2013_N		430_2013_N		430_2013_N		431_2013_N		431_2013_N		432_2013_N		432_2013_N		433_2013_N		433_2013_N		434_2013_N		434_2013_N		435_2013_N		435_2013_N		436_2013_N		436_2013_N		437_2013_N		437_2013_N		438_2013_N		438_2013_N		439_2013_N		439_2013_N		440_2013_N		440_2013_N		441_2013_N		441_2013_N		442_2013_N		442_2013_N		443_2013_N		443_2013_N		444_2013_N		444_2013_N		445_2013_N		445_2013_N		446_2013_N		446_2013_N		447_2013_N		447_2013_N		448_2013_N		448_2013_N		449_2013_N		449_2013_N		450_2013_N		450_2013_N		451_2013_N		451_2013_N		452_2013_N		452_2013_N		453_2013_N		453_2013_N		454_2013_N		454_2013_N		455_2013_N		455_2013_N		456_2013_N		456_2013_N		457_2013_N		457_2013_N		458_2013_N		458_2013_N		459_2013_N		459_2013_N		460_2013_N		460_2013_N		461_2013_N		461_2013_N		462_2013_N		462_2013_N		463_2013_N		463_2013_N		464_2013_N		464_2013_N		465_2013_N		465_2013_N		466_2013_N		466_2013_N		467_2013_N		467_2013_N		468_2013_N		468_2013_N		469_2013_N		469_2013_N		470_2013_N		470_2013_N		471_2013_N		471_2013_N		472_2013_N		472_2013_N		473_2013_N		473_2013_N		474_2013_N		474_2013_N		475_2013_N		475_2013_N		476_2013_N		476_2013_N		477_2013_N		477_2013_N		478_2013_N		478_2013_N		479_2013_N		479_2013_N		480_2013_N		480_2013_N		481_2013_N		481_2013_N		482_2013_N		482_2013_N		483_2013_N		483_2013_N		484_2013_N		484_2013_N		485_2013_N		485_2013_N		486_2013_N		486_2013_N		487_2013_N		487_2013_N		488_2013_N		488_2013_N		489_2013_N		489_2013_N		490_2013_N		490_2013_N		491_2013_N		491_2013_N		492_2013_N		492_2013_N		493_2013_N		493_2013_N		494_2013_N		494_2013_N		495_2013_N		495_2013_N		496_2013_N		496_2013_N		497_2013_N		497_2013_N		498_2013_N		498_2013_N		499_2013_N		499_2013_N		500_2013_N		500_2013_N		501_2013_N		501_2013_N		502_2013_N		502_2013_N		503_2013_N		503_2013_N		504_2013_N		504_2013_N		505_2013_N		505_2013_N		506_2013_N		506_2013_N		507_2013_N		507_2013_N		508_2013_N		508_2013_N		509_2013_N		509_2013_N		510_2013_N		510_2013_N		511_2013_N		511_2013_N		512_2013_N		512_2013_N		513_2013_N		513_2013_N		514_2013_N		514_2013_N		515_2013_N		515_2013_N		516_2013_N	
----------------	--	--------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--	------------	--

É feita uma validação do método, para saber se o modelo é adequado ou não, através da validação cruzada, como se pode observar na figura 28. São retiradas amostras uma a uma e de cada iteração é calculado um novo modelo. Com este, é feita a previsão da posição da amostra retirada.

Existiam ainda outras duas opções de validação em que consideram retiravam grupos de amostras aleatórias de forma a não fazer tantos modelos, estes modelos só são usados se não houver capacidade de processamento.

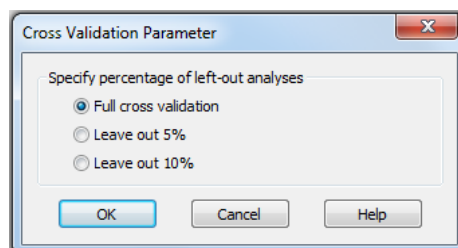


Fig.28 – Seleção do método de calibração a ser usado

É usado o critério em que apenas se considera 98% da variância que correspondem a apenas 78 CP, e não às cerca de 200 CP iniciais. As primeiras CP têm uma maior variância, mas não significa que contenham as diferenças esperadas (separação por regiões). Então, a escolha das CP a considerar é feita visualmente de acordo com a separação dos pontos. Neste caso, as componentes escolhidas foram as CP2 e CP8, como se observa pela figura 29.

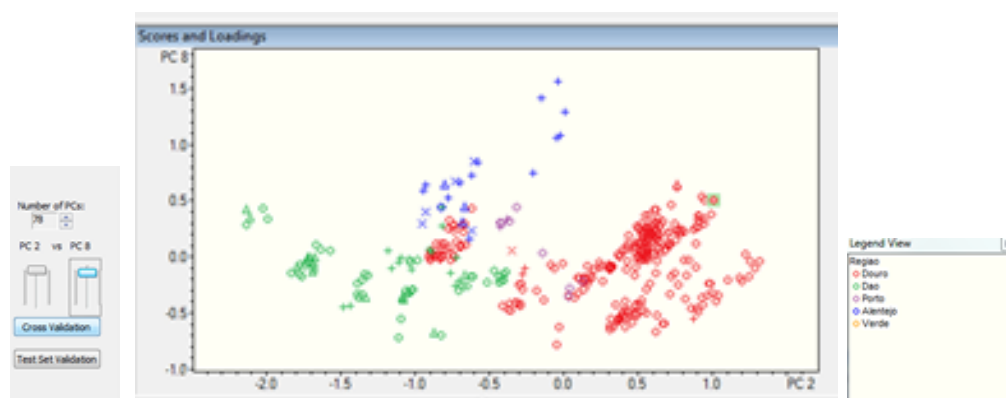


Fig.29 – Escolha das componentes principais de forma visual

Após a análise dos gráficos *influence*, *hotelling* e *distance to model*, é possível eliminar as amostras que se encontram fora das linhas, pois são aquelas que se encontram fora do

modelo. O segundo quadrante do gráfico *influence*, é aquele que contem as amostras fora do modelo com maior influência no resultado final, pois este é uma combinação dos outros dois gráficos já mencionados.

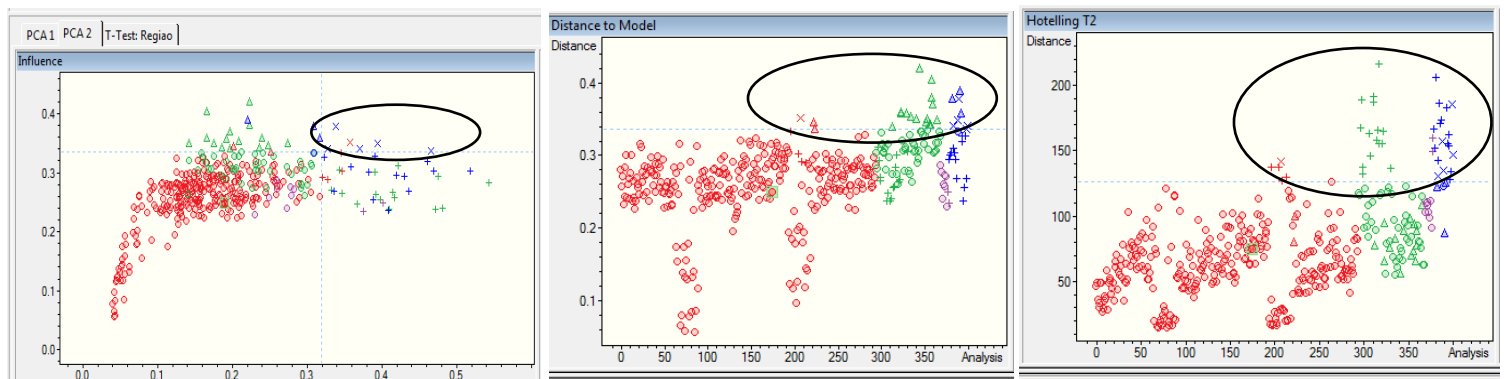


Fig. 30 – Análise da distância das amostras ao modelo

É ainda calculado um gráfico onde cada ponto corresponde a uma amostra (*scores*) e outro em que cada ponto corresponde a uma combinação entre o tempo e a razão *m/z* (*loadings*). Na representação gráfica das *loadings*, cada ponto representa uma variável e o plano definido por dois eixos correspondentes às componentes principais escolhidas, ou seja, a cada variável é associado um ponto cujas coordenadas são as correlações dessa variável com cada uma das componentes principais em causa. Na representação gráfica dos *scores*, cada ponto está associado um objeto, cujas coordenadas são os *scores* desse objeto para cada uma das componentes principais.

É possível observar através da figura 31, que apesar de existirem amostras com parâmetros semelhantes entre si, nota-se uma boa separação entre regiões. É de notar que o CP2 separa bem as amostras do Dão e do Douro, enquanto que o PC8 separa bem as amostras do Alentejo.

A variância mede o grau de diferenciação entre os elementos de um conjunto de dados, ou seja, quanto maior for a variância maior será a distinção entre as amostras. Como tal, os pontos centrais apresentam uma menor variância entre as amostras, enquanto nos pontos mais afastados é onde se encontram as maiores diferenças, havendo uma correspondência dos pontos entre os dois gráficos.

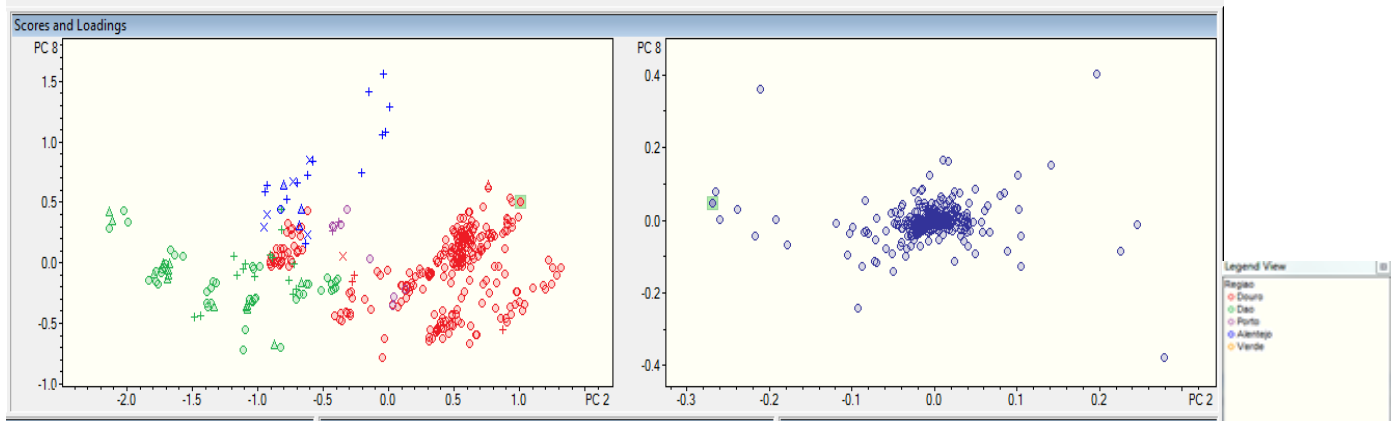


Fig. 31 – Semelhança das amostras entre si, e a distância de cada uma delas ao modelo (ionização em modo negativo das amostras de 2013)

Escolhendo duas amostras aleatórias que estejam mais afastadas do modelo, é possível observar que têm características que as diferem, dependendo da região. Através da figura 32 observa-se que o composto de  $m/z=127,004$  ao minuto 1,18 das amostras de 2013, ionizadas em modo negativo aparece com maior intensidade em amostras do Dão e o composto de  $m/z=125,000$  ao minuto 1,11 aparece com maior intensidade em amostras da região do vinho do Porto.

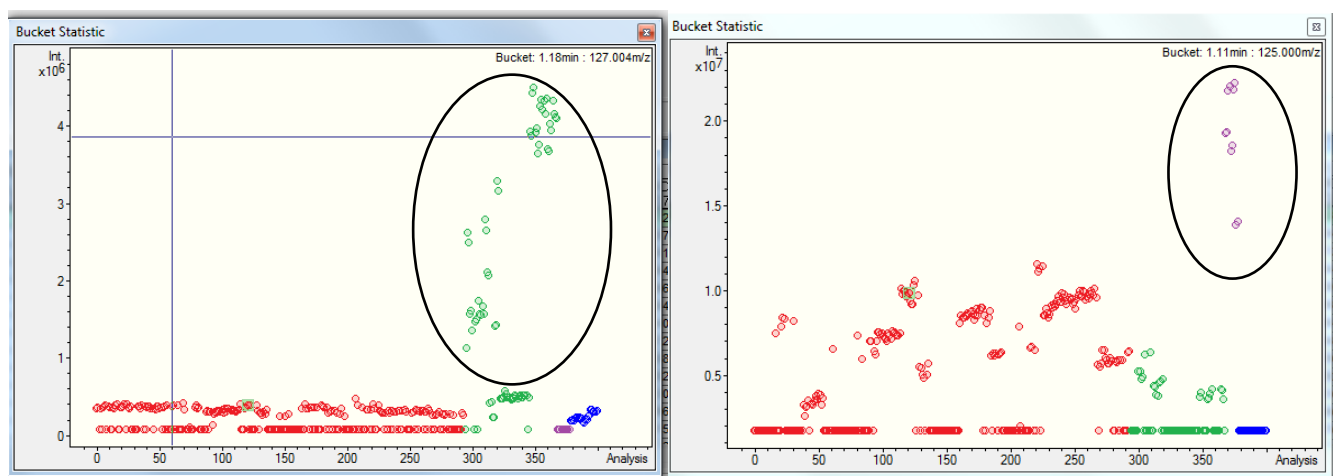


Fig. 32 – Variação da intensidade do *bucket* selecionado em todas as amostras



- **Diferenciação das castas em amostras de 2013 quando ionizadas em modo positivo**

Apenas foi tido em consideração a diferenciação das castas quando estas eram comuns em cinco ou mais amostras.

Analisando o figura 33, observa-se que as amostras que são constituídas pela monocasta TF(100%) e pelas castas TF(71%); TN(29%); TC(22%) não são bem separadas entre elas. Contudo, quando comparadas com as amostras constituídas pelas castas TF(49/57%); TC(22%); DIV(29/21%) já é possível observar uma boa diferenciação entre os grupos. Tendo em atenção à monocasta EN(100%), quando comparada com EN(60%); SE(40%), nota-se uma ótima separação entre as amostras.

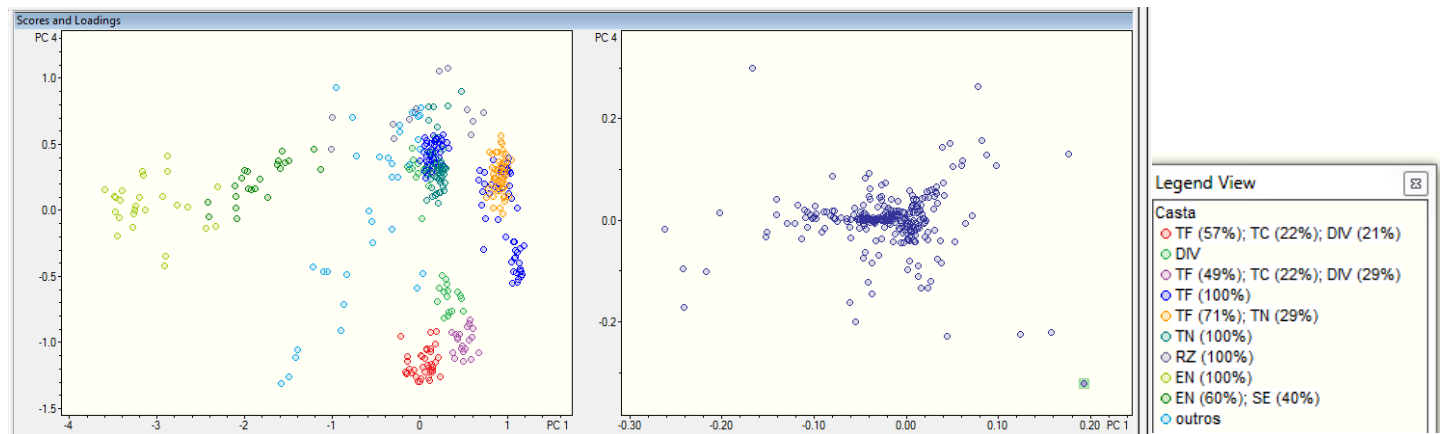


Fig. 33 – Diferenciação das castas em amostras de 2013

- **Comparação das amostras de 2013 quando ionizadas em modo negativo e em modo positivo**

Quando as amostras são ionizadas em modo positivo, a escolha das CP foi a 1 e a 6. O critério usado foi o mesmo, visualmente apresentam uma boa separação (observar o figura 34). A CP1 separa bem as amostras do Dão, Douro e Alentejo, enquanto que a CP6 separa bem as amostras do Porto. As amostras do Douro encontram-se muito bem agrupadas.

Quando as amostras são ionizadas em modo negativo, a escolha das CP foi a 2 e 8. A CP2 separa bem as amostras do Dão e Douro enquanto que a CP8 separa apenas as amostras do Alentejo. Ou seja, o vinho do Porto apresenta compostos diferentes dependendo do modo de ionização em que é analisado.

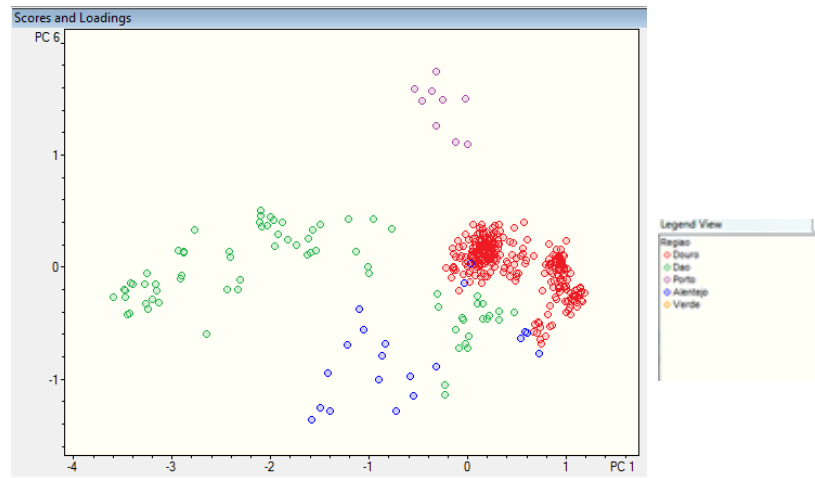
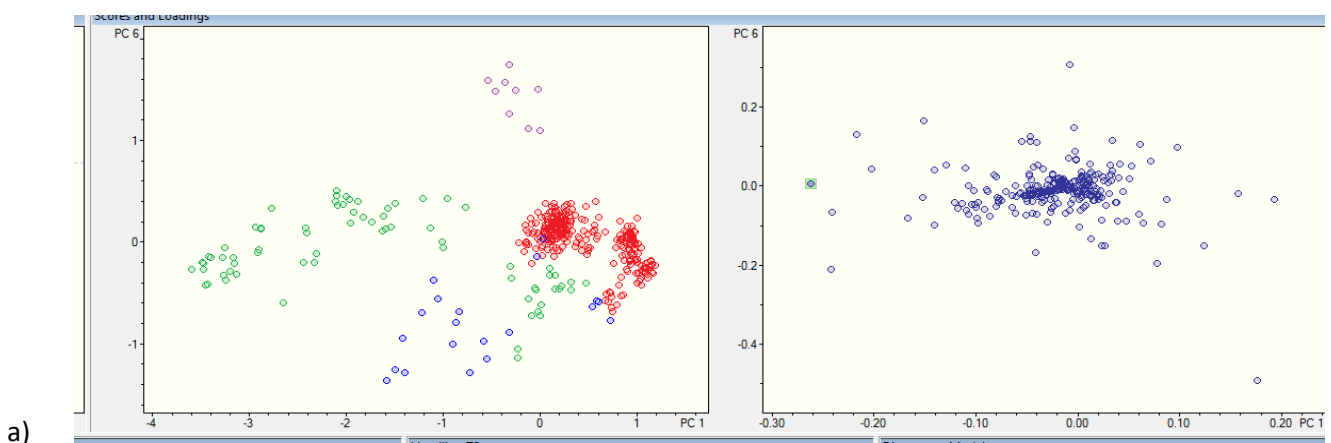


Fig. 34 – Escolha das componentes principais quando as amostras de 2013 foram ionizadas em modo positivo

Comparando o gráfico *loadings* das amostras do mesmo ano, mas com modos de ionização diferente, nota-se que em modo positivo existe um maior grau de diferenciação entre os elementos de um conjunto de dados.

Apesar das CP escolhidas entre os dois modos de ionização serem diferentes, nota-se que as amostras quando ionizadas em modo negativo apresentam uma melhor diferenciação entre regiões, daí existir uma maior variância das amostras no modelo (PCA).



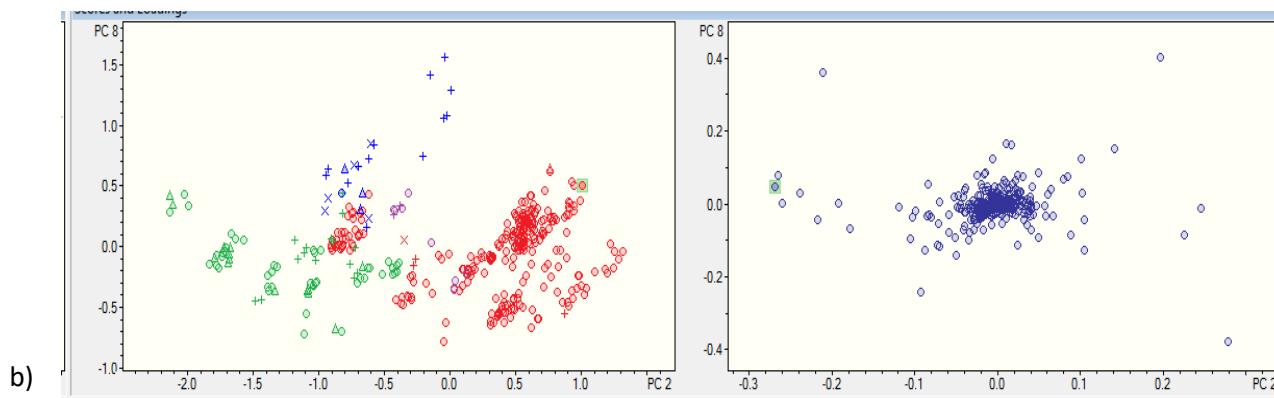


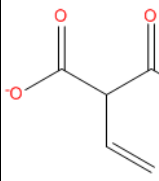
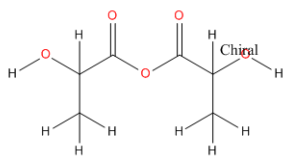
Fig. 35 – Comparação da variância dependendo do modo de ionização. a) ionização positiva; b) ionização negativa

Foi feita uma tabela dos *buckets* com maior variância (Ver anexo 1), de onde foram seleccionados alguns ao acaso de forma a tentar encontrar possíveis fórmulas de estrutura para os compostos de interesse.

Para cada *bucket* existe mais do que uma fórmula molecular, e cada fórmula molecular corresponde a mais do que um possível composto sendo um processo muito demorado e difícil de identificação. Na tabela 12, são apenas apresentados alguns exemplos.

Tabela 12 – Exemplos de possíveis estruturas para os *buckets* com maior variância

Modo de ionização	Tempo (min)	m/z	Fórmula molecular	Possíveis compostos	*Fórmula de estrutura
Positivo	1,20	118,085	C <sub>5</sub> H <sub>12</sub> NO <sub>2</sub>	L-Valina N,N-dimentil-L-alamina *Ácido 5-aminopentanóico	
	2,60	163,060	C <sub>6</sub> H <sub>10</sub> O <sub>5</sub>	3,6 Anidro-D-glucose Ácido 3-etilmálico *1,5-anidro-D-frutose	

Negativo	1,18	127,004	$C_5H_4O_4$	Ácido acónico 3-metoxi-2,5-furandiona *Vinilmalonato	
	3,14	161,045	$C_6H_{10}O_5$	Dietilpirocarbonato *Lactilactato	

- Diferenciação das amostras por ano 2013, 2014 e 2015 – modo positivo**

Após serem processadas as amostras em conjunto do ano de 2013 e 2014 (em modo de ionização positivo), observa-se pela figura 36, que no ano de 2013+2014 existe uma nítida separação entre as regiões, exceto nas amostras oriundas do Alentejo. Nota-se ainda que quando comparadas as amostras de 2013 com as amostras de 2013+2014 as amostras do Douro apresentam uma maior dispersão, enquanto que as do Dão mostram uma maior proximidade entre elas. Tal facto pode ser devido á nova criação de um modelo de PCA (análise de 2013+2014) em que as CPs podem não estar a diferenciar os mesmos parâmetros.

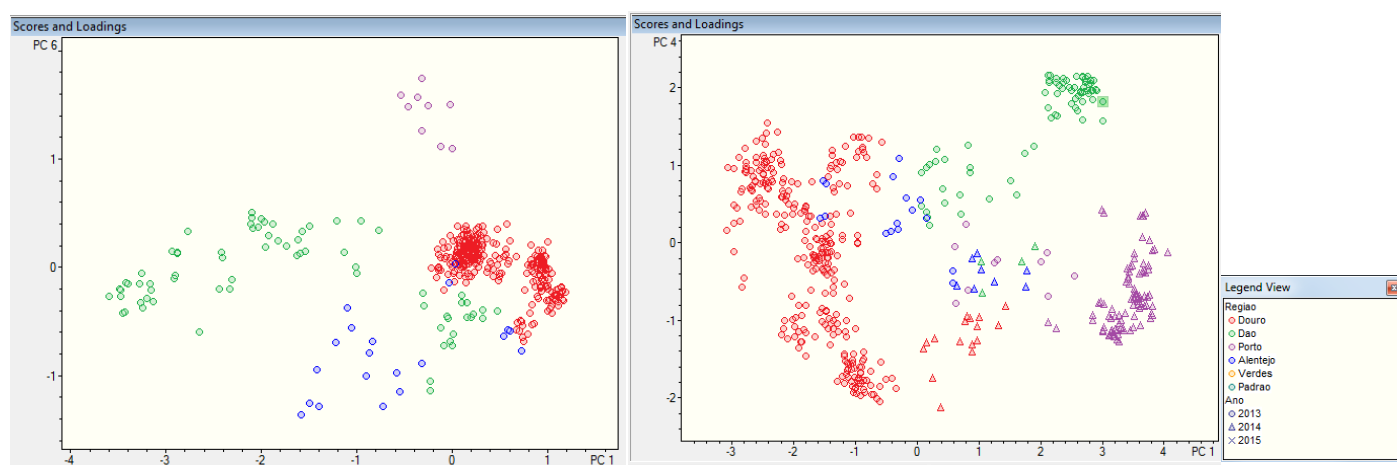


Fig. 36 – Diferenciação entre as amostras de 2013 e 2014

Quando analisadas todas as amostras dos diferentes anos (2013+2014+2015), observa-se uma ótima diferenciação entre as regiões todas elas agrupadas consoante as suas características, exceto da região do Alentejo que apresentam amostras com compostos comuns às do Douro. Contudo, existe uma boa separação entre o ano de 2013, mas o ano de 2014 apresenta uma fraca diferenciação com o ano de 2015.

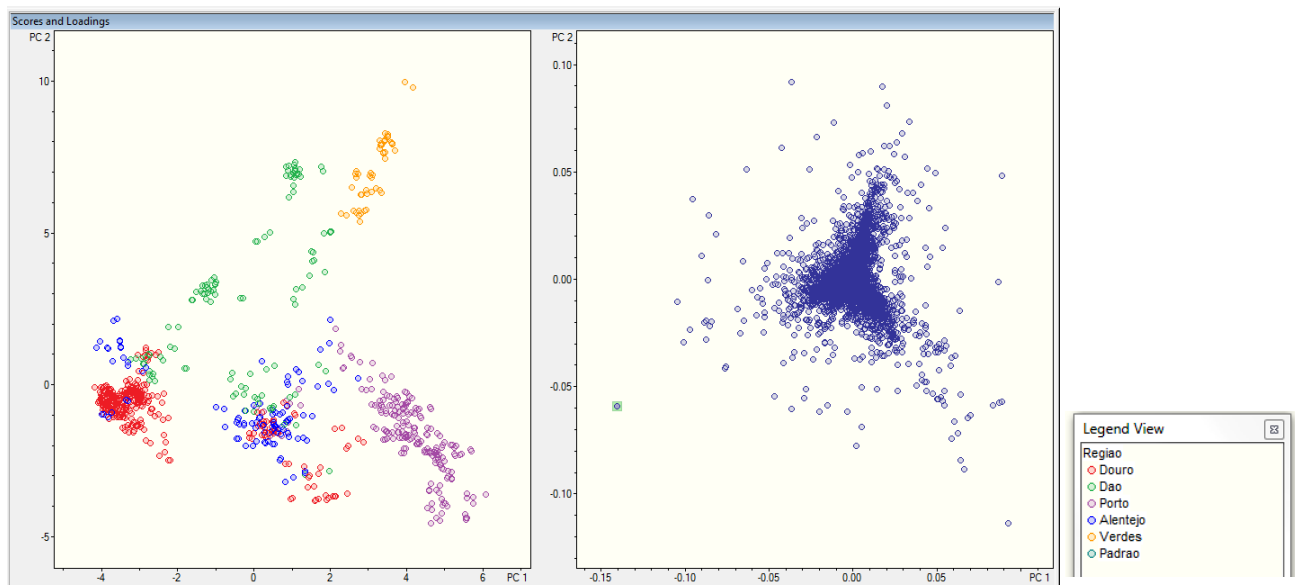


Fig. 37 – Análise das diferentes regiões das amostras de 2013, 2014 e 2015

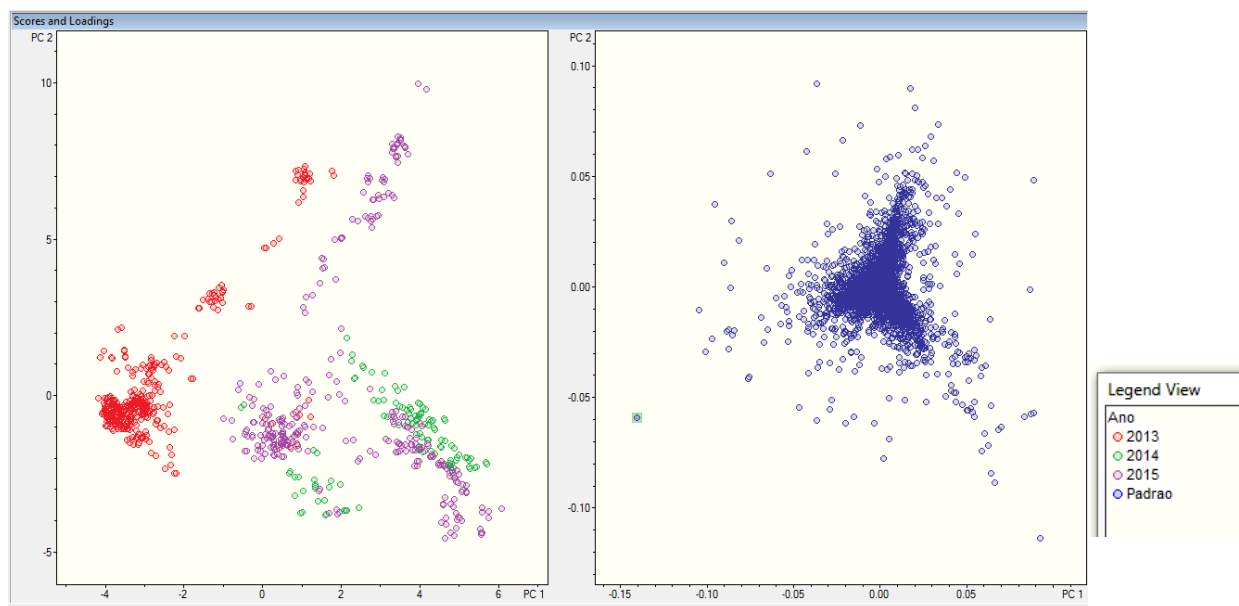


Fig. 38 – Análise dos diferentes anos das amostras 2013, 2014 e 2015

Após a análise das amostras de todos os anos pelos dois modos de ionização, o passo seguinte seria identificar todos os compostos com maior variância pela sua razão  $m/z$ . As amostras de interesse seriam novamente analisadas por LC-MS e posteriormente era feito o isolamento e fragmentação dos compostos de interesse e pela sobreposição dos espetros já era provável conseguir propor uma possível estrutura.

Para a identificação dos compostos era necessário recorrer a novos programas como o *Compound Crawler* onde é possível obter as fórmulas de estrutura e o *Smart fórmula* que é uma ferramenta do *DataAnalysis*, que fazem uma pesquisa de compostos nas bases de dados como o *MetFrag*, *kEGG*, *CheBI* e *ChemSpider*.

## Conclusões

O recurso a técnicas de análise sofisticadas, o conhecimento prático e empírico, revelam-se uma ferramenta eficaz na deteção de falsificações ou fraudes que possam ocorrer no fabrico do vinho. Esta temática tem atingido mundialmente as empresas produtoras de vinho, ocasionando perdas substanciais nos seus patrimónios.

Com a mesma velocidade de evolução das tecnologias, as fraudes e as formas com que elas acontecem, têm crescendo de forma considerável na sociedade atual. Os propósitos da realização das análises são diversos, podendo ter como intuito a deteção de falsificações, fraudes ou de determinar quantidades de alguns componentes de forma a verificar se estão dentro do regime legal imposto. A capacidade que uma empresa tem de detetar fraudes, revela-se uma ferramenta eficaz na descoberta de fatores capazes de direcionar todo o trabalho que é realizado, levando a empresa a um nível de qualidade e excelência no fabrico do vinho. Como tal, é necessário investigar mais sobre esta temática que tanto preocupa os produtores e os consumidores.

Para este trabalho foram propostos dois grandes objetivos, criar uma base de dados para a deteção de fraudes em vinhos e testá-la, recorrendo à análise por LC-MS e fazendo todo o seu tratamento estatístico por PCA. Nunca foi desenvolvido nenhum trabalho semelhante em Portugal e como tal este apresenta-se um ponto de partida para trabalhos futuros.

Foi feita uma boa amostragem pelas principais regiões em que a Sogrape produz vinho, Douro, Vinho do Porto, Dão, Alentejo e Verdes, com o intuito construir uma base de dados o mais robusta possível. Os parâmetros de análise e o procedimento experimental já estavam todos implementados devido a um trabalho semelhante já feito na empresa.

O LC-MS é um método de análise muito sensível, preciso e rápido o que permitiu, ao fim de compreender todo o seu funcionamento, analisar um elevado número de amostras. Após a análise por LC-MS todos os espectros de massa e cromatogramas foram calibrados e posteriormente analisados usando uma ferramenta estatística, o PCA.

Comparando as amostras do mesmo ano (2013) quando ionizadas de forma diferente (positiva e negativa), e apesar das CP escolhidas terem sido diferentes, nota-se uma boa separação entre as regiões. Quando a ionização foi feita em modo positivo, observou-se uma boa diferenciação e um bom agrupamento entre as amostras das diferentes regiões, independentemente das CP a considerar. Com a ionização negativa, as amostras da região do vinho do Porto não são separadas das outras. Foram selecionados *buckets* ao acaso, com o intuito de tentar encontrar possíveis fórmulas de estrutura para os compostos de interesse. Foram ainda diferenciadas as amostras pelas suas castas em que se observou diferenças no agrupamento dependendo da sua constituição.

Apesar de não ser possível analisar todas as amostras em modo negativo, em modo positivo obtiveram-se resultados do ano de 2013, 2014 e 2015. É de notar uma ótima separação entre todas as regiões, exceto a do Alentejo que em todas as amostras deve apresentar características iguais às do Douro, devido à sua sobreposição nos gráficos obtidos.

Visto que as condições do método a usar já estavam otimizadas, a grande limitação foi o software usado para o tratamento de dados, que devido ao elevado número de amostras não conseguiu processar todos os resultados de forma correta. Neste sentido, e após inúmeras tentativas de otimização de condições, verificou-se que o problema não era das condições escolhidas, mas sim da versão do software utilizado. Após a resolução deste problema, foi possível recolher algumas informações acerca das diferentes amostras em estudo.

Após a análise e tratamento estatístico de todas as amostras, ir-se-ia proceder à identificação e fragmentação dos compostos de interesse. Mais tarde, proceder-se-ia a uma nova injeção no LC-MS de amostras de vinho adulteradas de forma propositada e analisadas todas as diferenças, podendo verificar-se ou não a robustez da base de dados.



## Considerações finais e perspectivas futuras

A elevada pesquisa bibliográfica fez com que fossem adquiridos elevados conhecimentos dos métodos de análise ao vinho, do funcionamento do LC-MS e da importância da utilização de ferramentas estatísticas para o tratamento de resultados.

Foi possível realizar uma ótima amostragem pelas principais regiões de produção da Sogrape Vinhos SA, permitindo uma elevada robustez na base de dados a ser criada. Foram analisadas 414 amostras oriundas de 5 regiões vinícolas diferentes.

Verificou-se uma boa separação entre as amostras tendo em conta o seu ano e região, e no caso de 2013 até pela diferenciação das suas castas.

Apesar de não terem sido processadas todas as amostras, todas as condições estão otimizadas para a continuação deste desafiante trabalho que tanto poderá ajudar os produtores e consumidores de vinho.

## Referências Bibliográficas

- [1] Abdi, H. e Williams, L. J. (2010). Principal component analysis, John Wiley & Sons, Inc.,2: pp. 433-459.
- [2] Afonso,J. (2006). Os componentes do Vinho (parte 1), revista ADEGA, edição 7, Acedido em 17/12/2015, em [http://revistaadega.uol.com.br/artigo/os-componentes-do-vinho-parte-1\\_5934.html](http://revistaadega.uol.com.br/artigo/os-componentes-do-vinho-parte-1_5934.html)
- [3] Alfana Wines. Acedido em 30/11/2015, em <http://www.alfamawines.pt/regioes-vincolas.html>.
- [4] Alpuim, J. (1997). Aprendendo a Química do Vinho, Química, 65: pp. 13-27.
- [5] Alves, M. (1992). Caracterização Química de 4 castas produtoras de Vinho do Porto durante a maturação. Tese de Doutoramento apresentada á Faculdade de Engenharia da Universidade do Porto.
- [7] Andrade, M. (2008). Caracterização e distinção analítica e quimiométrica de vinhos elaborados no vale de S. Francisco e no Rio Grande do Sul, Dissertação de Mestrado apresentada ao Departamento de Química da Universidade Federal de Pernambuco.
- [8] Araújo, R. (2008). Validação de metodologia de análise de pesticidas agrícolas em águas por LC/MS. Projeto de Mestrado Integrado em Engenharia Química. Faculdade de Engenharia da Universidade do Porto, Porto.
- [9] Barak, J. A. e Kennedy, J. A. (2013). HPLC Retention Thermodynamics of Grape and Wine Tannins, Journal of Agricultural and Food Chemistry, American Chemical Society. pp. 4270-4277.
- [10] Borràsa, E.; Ferréb, J.; Boquéb, R.; Mestresa, M.; Aceña, L. e Bustoa, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment – A review. Analytica Chimica Acta, 891: pp. 1–14
- [11] Bro, R. e Smilde, A.K. (2014). Principal component analysis, Journal of The Royal Society of Chemistry, Anal. Methods, 6: pp. 2812–2831.
- [12] Cadima, J.F. C. L. Redução de dimensionalidade através duma análise em componentes principais: um critério para o número de componentes principais a reter, Departamento de Matemática - Instituto Superior de Agronomia, pp.1-9.

- [13] Cordella, C. B.Y. (2012). PCA: The Basic Building Block of Chemometrics, Analytical Chemistry, InTech, Chapter 1, pp. 3-35.
- [14] Cruz, L.; Mateus, N. e de Freitas, V. (2012). Identification by mass spectrometry of new compounds arising from the reactions involving malvidin-3-glucoside-(O)-catechin, catechin and malvidin-3-glucoside, Wiley Online Library. pp. 2123–2130.
- [15] Fernandes, S. (2010). Desenvolvimento de um método de análise de Pesticidas na Indústria Vitivinícola, Dissertação de Mestrado em Química Área de Especialização em Técnicas de Caracterização e Análise Química apresentada à Universidade do Minho, pp. 19-25, 39-43.
- [16] Figueiredo-González, M.; Cancho-Grande, B.; Simal-Gándara, J.; Teixeira, N.; Mateus, N. e de Freitas, V. (2014). The phenolic chemistry and spectrochemistry of red sweet wine-making and oak-aging, Elsevier, Food Chemistry, 152: pp. 522–530.
- [17] Gas Chromatography - Tandem Techniques - The Basic Principles of Mass Spectroscopy. Acedido em 17/12/2015, em [http://www.chromatography-online.org/Mass-Spectroscopy/rs\\_13\\_34.php](http://www.chromatography-online.org/Mass-Spectroscopy/rs_13_34.php)
- [18] Gordillo, B.; Rodríguez-Pulido, F.J.; Mateus, N.; Escudero-Gilete, M.L.; González-Miret, M.L.; Heredia, F.J. e de Freitas, V. (2012). Application of LC–MS and tristimulus colorimetry to assess the ageing aptitude of Syrah wine in the Condado de Huelva D.O. (Spain), a typical warm climate region, Elsevier, Analytica Chimica Acta, pp. 162–171.
- [19] Han, S.; Wang, J.; Liu, Y.; Zhou, Q.; Lu M. e Zhou S. (2014). Determining Wine Authenticity: A Metabolomics Analysis Using UHPLC ESI/Q-TOF MS. Agilent Technologies, USA.
- [20] Hermann, G. (2013). Analysis of (poly)phenols in red wine via LC-MS. Masterarbeit in Biotechnologie. Universität für Bodenkultur Wien University of Natural Resources and Life Sciences, Vienna.
- [21] Ho, C.S.; Lam, C.W.K.; Chan, M.H.M.; Cheung, R.C.K.; Law, L.K.; Lit, L.C.W.; Ng, K.F.; Suen, M.W.M. e Tai, H.L. (2003). Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications, Clin Biochem Ver, 24: pp. 3-12.
- [22] Instituto da Vinha e do Vinho – Ministério da Agricultura, Castas. Acedido em 30/11/2015, em <http://www.ivv.min-agricultura.pt/np4/np4/153.html>.
- [23] Instituto da Vinha e do Vinho – Ministério da Agricultura, Regiões. Acedido em 30/11/2015, em <http://www.ivv.min-agricultura.pt/np4/regioes>.

- [24] Kang, J.S. (2012). Principles and applications of LC-MS/MS for the quantitative bioanalysis of analytes in various biological samples. InTech. Rijeka, Croácia. pp. 441-492
- [25] Martim, A. (2004). L'analyse de données. Polycopié de cours ensieta – 1463: pp. 23-39
- [26] McMaster, M. (2005). LC/MS A Pratical User's Guide. John Wiley&Sons, Inc., Hoboken, New Jersey. pp.51-81
- [27] Melo, F. (2011). Utilização de técnicas microbiológicas na avaliação da eficiência de um sistema de HACCP a nível de adega, Dissertação de Mestrado Em Microbiologia Aplicada apresentada à Faculdade de Ciências da Universidade de Lisboa.
- [28] Moreira, A.C. (2007). Comparação da Análise de Componentes Principais e da CATPCA na Avaliação da Satisfação do Passageiro de uma Transportadora Aérea. Investigação Operacional, 27: pp. 165-178
- [29] O Portal do Vinho Português – Infovini, Castas. Acedido em 2/12/2015, em <http://www.infovini.com/pagina.php?codNode=18017#tab0>
- [30] O Portal do Vinho Português – Infovini, Regiões. Acedido em 30/11/2015, em <http://www.infovini.com/pagina.php?codNode=18012>.
- [31] Pinheiro, G.; Cunha, L. (2003). A importância da auditoria na deteção de fraudes, Contab. Vista & Rev. Belo Horizonte, 14: pp 3-4.
- [32] Rosado, A. (2013). Evolução de Parâmetros Físicos, Químicos e Controlo Microbiológico em Vinhos Brancos e Tintos da Adega Cooperativa de Palmela, Dissertação para obtenção do Grau de Mestre em Tecnologia e Segurança Alimentar apresentada à Universidade Nova de Lisboa, pp. 1-3, 12-17.
- [33] Silva, T. E. V.; Vasconcelos, F. H. L.; Mota, J. C. M.; Almeida A.L. F. e Andriola, W. B. (2013). Ferramentas de Processamento e Análise de Informação Aplicadas ao Ensino de Engenharia. XXXI Simpósio Brasileiro de Telecomunicações. Fortaleza.
- [34] Simonato, B.; Mainente, F.; Tolin, S. e Pasini, G. (2011). Immunochemical and Mass Spectrometry Detection of Residual Proteins in Gluten Fined Red Wine. Journal of Agricultural and Food Chemistry, American Chemical Society. pp. 3101-3110.
- [35] Sogrape Vinhos, S.A. Acedido em 9/10/2015, em <http://www.sograpevinhos.com/sogrape/historia>.

- [36] Vagante, C. (2012). Efeitos do consumo de vinho na saúde humana: Aspectos positivos e negativos, Dissertação para obtenção do Grau de Mestre em Tecnologia e Segurança Alimentar apresentada à Universidade Nova de Lisboa, pp. 1-28.
- [37] Valle, I. (2012). Aplicação de ferramentas estatísticas e de métodos quimiométricos clássicos no tratamento da informação multivariada relativa à caracterização físico-química de vinhos. Dissertação de Mestrado em Química – Controlo da Qualidade e Ambiente. Departamento de Química da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Coimbra. pp. 98 – 111.
- [38] van Galen, P. M. (2005). Mass Spectrometry - A guide for novel users, Organic Chemistry Department - Nijmegen University, pp. 2-12, 19,27-28.
- [39] Vilela, A.; Monteiro, B.; e Correia, E. (2015). Perfil sensorial de vinhos do porto: análise em componentes principais categórica, uma abordagem para o tratamento de dados sensoriais, Ciência Téc. Vitiv., 30: pp. 1-8.
- [40] Wines of Portugal: a World of difference. Acedido em 30/11/2015, em <http://www.winesofportugal.info/pagina.php?codNode=18012&market=1>.

## ANEXOS

### Anexo 1 – Tabela dos *buckets* com maior variância do ano de 2013

Modo de ionização positivo		Modo de ionização negativo	
Tempo (min)	Massa (Da)	Tempo (min)	Massa (Da)
0,76	216,95	0,80	272,96
0,8	177,01	0,83	288,94
0,85	95,06	0,97	537,06
0,86	112,09	0,98	96,96
1,07	667,23	1,00	367,05
1,13	363,07	1,12	125,00
1,24	72,08	1,16	189,00
1,24	118,08	1,17	193,03
1,62	87,04	1,17	127,00
1,66	381,08	1,19	161,04
2,07	283,13	1,21	173,04
2,1	219,13	1,23	134,02
2,16	247,13	1,74	159,03
2,16	260,16	2,85	200,05
2,38	116,07	2,95	339,13
2,45	87,04	3,02	177,44
2,51	87,04	3,07	616,11
2,58	116,07	3,13	161,04
2,62	163,04	3,19	299,08
2,65	340,11	3,33	219,05
2,82	86,09	3,36	155,00
2,95	318,17	3,39	173,01
3,08	279,13	3,60	161,05
3,11	149,06	3,64	179,03
3,14	86,09	3,66	366,12
3,14	86,09	3,90	131,07
3,16	132,1	4,18	147,04
3,18	158,08	5,32	327,22
3,26	229,15	5,54	329,23
3,26	229,15		

3,55	199,06
3,88	101,02
3,99	211,14
3,99	259,2
4,05	291,17
4,36	127
4,38	160,13
4,39	273,22
4,72	130,12
5,45	353,23